Volume 12
Number 1 *Numbers 1, 2, & 3*

Article 1

6-2015

TME Volume 12, Numbers 1, 2, and 3

Editorial: The Economics of Risk

Bharath Sriraman[1]

University of Montana

Vol.12 of The Mathematics Enthusiast addresses the notion of risk from a variety of viewpoints, mathematical or otherwise. This editorial examines risk, starting from a historical perspective to frame current educational policies that have been influenced by ensuing economic principles. Risk has historically been associated with numerous cultures in the context of dice games or gambling (Sriraman & Lee, 2014) .Risk was first empiricized in the financial sector during the advent of western colonialism in late 15th century when ships made voyages to the East and investors needed assurance about the risk associated in relation to the payout from a successful voyage. One sees the birth of maritime insurance during this period to address the needs of the mercantile sector which invested in risky shipping ventures. Another area in which risk was empiricized and calculated was in the domain of life insurance. It is well known among historians of statistics, that Edmond Halley in the 17th century constructed a life table to demonstrate how premiums can be calculated as a function of age. Even though the death of an individual was unpredictable, data on patterns of longevity for groups became a basis of study whereby risk could be more accurately assessed for the purpose of calculating premiums or pensions (Halley, 1693).

The economic dimension of risk in the 21st century is ubiquitous with existence in the developed world and forms the basis of neo-liberal economies and their associated markets. This ranges from auto-insurance premiums to investments made by governments for research, development and education. Educational reform in the U.S has been tied to risk as seen in the 1983 report from the National Commission on Excellence in Education under the auspices of the Reagan administration. The report entitled "*A Nation at Risk*" is often cited in the literature when addressing the failure of public schools and the need for reform in public education. It became somewhat of a pre-cursor for the private sector to intervene in public education since tax dollars spent on schools needed to be accounted for in terms of deliverable and measurable outcomes. Subsequent legislation in the form of "*No Child Left Behind*" (NCLB) enacted in 2001 held public schools accountable on the basis of student performance. The goal of this legislation was not only to reform public education but also to systematically close public schools that placed "students at risk" through probationary measures on the basis of student performance outcomes on standardized tests. By holding public schools and teachers responsible for inadequately addressing student needs, the onus of reform could be shifted to charter and private schools. One can argue that the increased role of the federal government in public education in the U.S has resulted in the marketization of public education by allowing the private sector to fill in the gap left by public schools. Funding initiatives such as the "*Race to the Top*" which were part of the *American Recovery and Reinvestment Act of 2009* furthered market economy goals of "competition" to put into place a reward system for states and schools that adhered to educational policies for schools anchored in performance based assessments and other measurable deliverables.

[1] sriramanb@mso.umt.edu

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 1 – 2

A consequence of this has been a further increase in charter schools in states like Illinois and a demise of the public school system. In a neo-liberal economy such as the U.S the presence of free public schools can be viewed as an anomaly or even a thorn in market principles such as competition and performance based incentives. However the risk of sacrificing equitable public education through reform anchored on the principles of neo-liberalism have numerous consequences. Ambrose (2012) argued that

> *The "school reforms" promoted by neoliberal ideologues are punitive and lead to further socioeconomic segregation of students. Those from impoverished backgrounds must languish in schools that are being punished for poor results on superficial standardized tests while those from privileged backgrounds can enjoy more creative, engaging, and challenging forms of…[l]earning.*

On a much larger scale, assessments such as PISA (*Program for International Student Assessment*) reveal socio-economic inequalities within countries with students of lower socio-economic status (SES) performing poorly in comparison to their higher socio-economic status peers. This is particularly evident in countries with economies anchored in a neo-liberal ideology, and increasingly evident in countries that have adopted these principles. Aspects of international assessments such as PISA in terms of students at risk due to low SES have been the topic of much discussion within educational circles (Štrajn, 2014). While PISA reveals inequities of scores as a function of SES in countries such as the U.S, it is often viewed as an instrument of the Organization for Economic Co-operation and Development(OECD), an economic organization of mostly rich and developed countries, to influence national educational policies. The paradox for the U.S lies in the fact that left leaning educational theorists criticize it as an instrument of neo-liberal ideology from an economic organization, whereas conservatives view it as an instrument of social democratic principles imported from Europe! This begs the question, what does happen to students of low SES whose lives are governed by a larger economy of neo-liberal risk?

References

Ambrose, D. (2012). Dogmatic neoclassical economics and neoliberal ideology suppressing talent development in mathematics: Implications for teacher education. In L. J. Jacobsen, J. Mistele, & B. Sriraman (Eds.), *Mathematics teacher education in the public interest: Equity and social justice*. Charlotte, NC: Information Age Publishing

Halley, E. (1693). An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw: An attempt to ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal Society of London*, vol. 17, 192-206.

Sriraman, B., & Lee, K.H. (2014). The humanistic dimensions of probability. In E.J. Chernoff and B. Sriraman (Eds) *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 117-119). Springer Science and Business, Berlin/Heidelberg

Štrajn, D. (2014) The PISA Syndrome: Can we Imagine Education without Comparative Testing? In M. Štraus (Ed). *Šolsko polje: Evidence from the PISA Studyon Educational Quality in Slovenia and Other Countries* (pp. 13-27)*.* Revija za teorijo in raziskave vzgoje in izobraževanja, Letnik XXV, Slovenia.

Guest Editorial: Risk – Mathematical or Otherwise

Egan J Chernoff (@MatthewMaddux)
University of Saskatchewan, Canada

‘Risk’ is an established domain of research in various academic disciplines (e.g., mathematics, statistics, probability, economics, engineering, political science, business, earth science, health science, computer science, psychology, sociology, and law). At present, risk, as a domain of research, is not similarly established in the field of mathematics education. However, there is established research on risk from members of the mathematics education community (e.g., Dave Pratt, Manfred Borovcnik, Ramesh Kapadia and others). Further, the established research on risk from the fields of psychology (e.g., Daniel Kahneman, Amos Tversky, Paul Slovic, Gerd Gigerenzer and others) and the popularization of mathematics and statistics (e.g., David Spiegelhalter and others) is consistently found in the relevant mathematics education research literature. As such, one could argue that risk, as a domain of research in the field of mathematics education, is starting to take shape.

The signs are there for risk to become a major topic of research in mathematics education. For example, there are themed issues of prominent mathematics education journals dedicated to “Probability and Reasoning about Data and Risk” (Biehler & Pratt, 2012), keynotes and plenaries on risk at prominent international conferences (e.g., 9th International Conference on Teaching Statistics) and, of course, foolish prognostications (Chernoff & Sriraman, 2014). Lest we forget, however, risk can be defined in numerous ways and is an interdisciplinary domain of research. As such, any publication attempting to help further cement risk as an established domain of research in the field of mathematics education and beyond would have to have rather unique features; which would include, but would not be restricted to, such features as: the open accessibility of material, a reputation of inviting articles that present original work on a wide range of topics and from various academic disciplines. Alternatively stated, a special issue on ‘Risk – Mathematical or Otherwise’ befits *The Mathematics Enthusiast* and *The Mathematics Enthusiast* befits a special issue on ‘Risk – Mathematical or Otherwise’

This special issue aspires not only to further investigate established research threads pertaining to risk, but, also, to identify and investigate, that is, extend the purview of risk research in the field of mathematics education and beyond. To achieve this lofty goal, over 40 individuals from various academic disciplines (e.g., mathematics, statistics, psychology, popularization, education and cognitive science) have contributed 26 articles, which, collectively, comprise the broad-based, interdisciplinary special issue on risk, mathematical or otherwise, that I had hoped to achieve when I was invited to guest edit this special issue. I wish to thank those individuals, here, for helping my plan come to fruition.

References

Biehler, R., & Pratt, D. (Eds.) (2012). Probability in reasoning about data and risk [special issue]. *ZDM—The International Journal on Mathematics Education, 44*(7), 819–952.

Chernoff, E. J., & Sriraman, B. (2014). Commentary on *Probabilistic Thinking: Presenting Plural Perspectives*. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 721-728). Berlin/Heidelberg: Springer Science.

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 3

What Can Education Learn from Real-World Communication of Risk and Uncertainty?

David Spiegelhalter
University of Cambridge, England
Jenny Gage
University of Cambridge, England

Abstract: Probability is a difficult topic to teach, not least because it is rather unclear what it actually means. Modern risk communication has tackled general public incomprehension of probability statements by using the metaphor of 'expected frequencies' – for example, "of 100 people like you, we would expect 10 to have a heart attack or stroke in the next 10 years." We show how these ideas can be taken into the classroom as the basis for teaching probability, using frequency tree diagrams as the fundamental representation. Empirical frequency trees can be used to summarise a series of classroom experiments, and then expected frequency trees naturally provide a basis for deriving the rules of probability, and make complex conditional probability calculations reasonably straightforward.

Keywords: natural frequencies, expected frequency, frequency trees.

Introduction

Risk-communication is a hot topic, whether it concerns the benefits and harms of screening or the chance of a catastrophic earthquake. It is challenging to explain both unpredictability and uncertain knowledge to the public, and yet these are also essential elements in education in probability and statistics. I shall argue that current approaches in communicating risk and uncertainty can contribute substantially to educational practice.

In particular, Gerd Gigerenzer's recommendation for 'natural frequencies' – whole-number outcomes starting from a defined population of cases - can be adapted to teaching probability based on a natural sequence of stages: empirical multiple narratives from experimentation represented as 2-way tables and frequency trees, to expected outcomes in multiple future experiments, and finally to probability trees. Issues of relative and absolute risk continually arise in topical stories, and representations that make these transparent are as relevant in the classroom as in the news.

I will first confront a fundamental question:

Why do people find probability and statistics unintuitive and difficult?

I've been working in this area for around 35 years, and after all this time have finally arrived at an answer.

Because probability and statistics are unintuitive and difficult.

I think it is important to remember this, and have sympathy with the struggle for comprehension. When someone asks me a school-level probability question, it's not like some algebra when I know the steps and just have to slog through to the answer. Probability questions always require some careful thought, and usually checking via at least two different solution methods. And the wording is so crucial (and often unclear).

I think that part of this difficulty is that it is unclear (to me at least) what probability actually *is*. It seems to be 'virtual' number – it is not directly measurable as is distance or weight. In fact I come from the subjectivist Bayesian school, founded by Italian statistician Bruno de Finetti, that holds that *probability does not exist* as a property of the external world (except maybe at a sub-atomic level), and so any number we use is a construct based on our current (imperfect) knowledge.

So any description of probability is a metaphor, and we can expect that different ways of describing probability may be interpreted in very different ways.

Communicating Probabilities

I often ask medical audiences a question like:

If you had to tell a person that around 20% of cases with their condition would suffer a certain bad outcome within 10 years, would you use the terms -

1. *20% chance*
2. *0.2 probability*
3. *1 in 5 chance*
4. *20 out of 100 cases like you*
5. *None of the above*
6. *Wouldn't use numbers*
7. *Depends on person*

They tend to prefer option 4, which fortunately turns out to the one supported by psychological research.

In particular, using many '1 in X' statements is not recommended. In a recent population survey by telephone, Galesic and Garcia-Retamero (2010) asked

Which of the following numbers represents the biggest risk of getting a disease:

1 in 100, 1 in 1000, 1 in 10 ?

In Germany, 28% of responses were incorrect, and in US 25% got it wrong. The crucial issue is using larger numbers to communicate smaller risks, so a difficult inversion must be done: this is one reason why flood risk maps expressed in terms of '1 in 100 year event' etc are difficult to read and potentially misleading.

But even if using whole frequencies to express risk, such as 20 out of 100, it must be kept in mind that 200 out of 1000 tends to appear bigger (Denes-Raj, Epstein & Cole, 1995) – this is known as 'ratio bias' in which focus is on the size of the numerator. The extreme version, in which the denominator is ignored completely, is known as "denominator neglect": the media does this every time they concentrate on a single accident without, for example, mentioning the millions of children that go to school safely each day. Remember: Newspapers like Narratives using Numerators.

Absolute and Relative Risks

A US direct-to-consumer advert for a statin declares in large font '36% reduction' in the risk of heart attack. In very much smaller font it clarifies that this is a reduction, in absolute percentage terms, from 3% to 2% over 5 years. So 100 such people would have to take the drug every day for 5 years to prevent one heart attack. This does not sound so impressive.

Gerd Gigerenzer has popularised the idea of "natural frequencies": translating probability problems into expectations for, say, 100 or 1000 people (Gigerenzer & Edwards, 2003). Take a recent headline saying that eating 50 grams of processed meat each day (eg a bacon sandwich) was associated with a 20% increased risk of pancreatic cancer (Willey, 2012). It turns out that this very serious disease fortunately affects only 1 in 80 people. So we want to calculate a 20% increase on a 1-in-80 chance, which is tricky to do.

However, if we imagine 400 people who have a nice healthy breakfast, we can easily calculate that we would expect 5 to get pancreatic cancer. If 400 other people all stuff themselves with a greasy bacon sandwich every day of their lives, this 5 would increase by 20% - to 6. Again, this 1-in-400 increase does not seem so important. Note the trick in identifying 400 as the denominator that will just give an impact of one case.

Expected Frequency Trees

In the previous example, the denominator of 400 was deliberately chosen so that the crucial difference comprised a single person. A similar exercise was conducted for the recent revision of the advice leaflets for breast cancer screening in the UK. I was on the panel that worked on this controversial topic, using evidence from a review of an independent panel (Breast screening benefits and harms, 2010), and with the approach that the leaflets would present the potential benefits and harms of screening, but would not make an explicit recommendation.

We drew up the 'expected frequency tree' shown in Figure 1, comparing the expected experience of 200 women with and without screening.

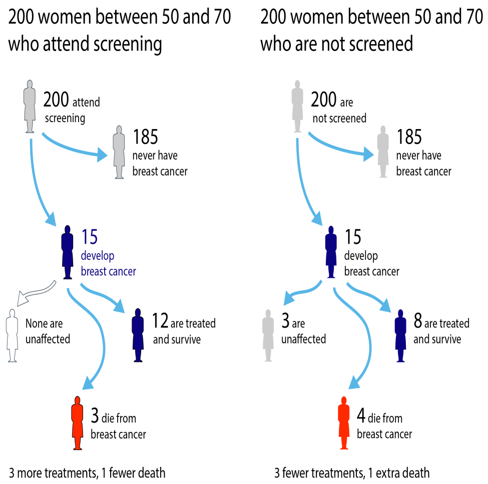


Figure 1. Expectations for 200 women attending or not attending breast screening every 3 years between the ages of 50 and 70

A website (NHS breast screening, n.d.) that incorporated a similar infographic recently won a 2014 UK Association of Medical Research Charities Science Communication Award.

Our 'Manifesto' for Teaching Probability

With Dr Jenny Gage of the Millennium Mathematics Project in Cambridge, we have developed a 'manifesto' for teaching probability that exploits the ideas of narratives, multiple representations, natural frequencies, expectation trees and so on. This can all be viewed on the Nrich website (Great expectations: probability through problems, n.d.).

Put simply, the stages are:

•Start with a problem (necessarily simplified to some extent)

•Model physically (using simple equipment, such as a die with different coloured faces or small coloured cubes)

•Do experiments (in groups, recording outcomes)

•Pool empirical data to represent multiple 'narratives' as

  •2 x 2 tables

  •Frequency tree

  •Venn diagram

•Spot patterns (classroom discussion to see what hypotheses emerge from the data, rather than asking students to predict in advance what will happen when they have no experience on which to base an opinion)

•If available, conduct large numbers of experiments using animations

•Ask questions about *expectation* in multiple future experiments

•Construct expected frequency tree

•Change to fractions on branches to get probability tree

Note that probability comes in as the *final* step.

Example: The Dog Ate My Homework!

A certain teacher, Mr L I Detector, claims he can tell when students are lying about their homework. This is true. Unfortunately, he also accuses some students who are telling the truth. So what are the chances that someone will be wrongly accused?

Note the deliberate ambiguity of the final question – different possible interpretations will arise naturally.

We use dice and 'lego-like' cubes to carry out multiple experiments of situations in which a student claims that their homework has been eaten by a dog. The first throw identifies whether a student is telling the truth or not: a 'six' means they are lying, and they get a red cube, otherwise a blue. Then the 'honest' ones throw again to decide the teacher's opinion: if a 'one' then he accuses them (yellow cube added to blue cube), otherwise a green cube (added to blue cube). The dishonest ones always get accused, in this version of the story, and so get a yellow cube to add to their red cube.

The pairs of coloured cubes summarise independent narratives and can be accumulated as a 2x2 table – this idea was derived from previous experiments by Martignon and Kurz-Milcke (2006) and Kurz-Milcke, Gigerenzer and Martignon (2008).

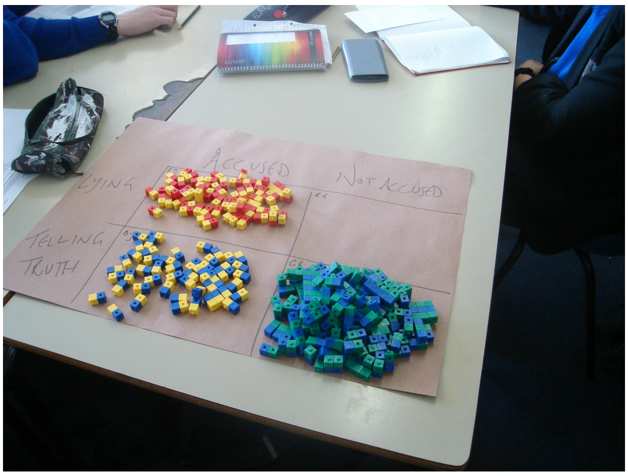


Figure 2. 2x2 table arising from multiple experiments of the 'dog' question

The data can then also be represented as an empirical frequency tree as in Figure 3.

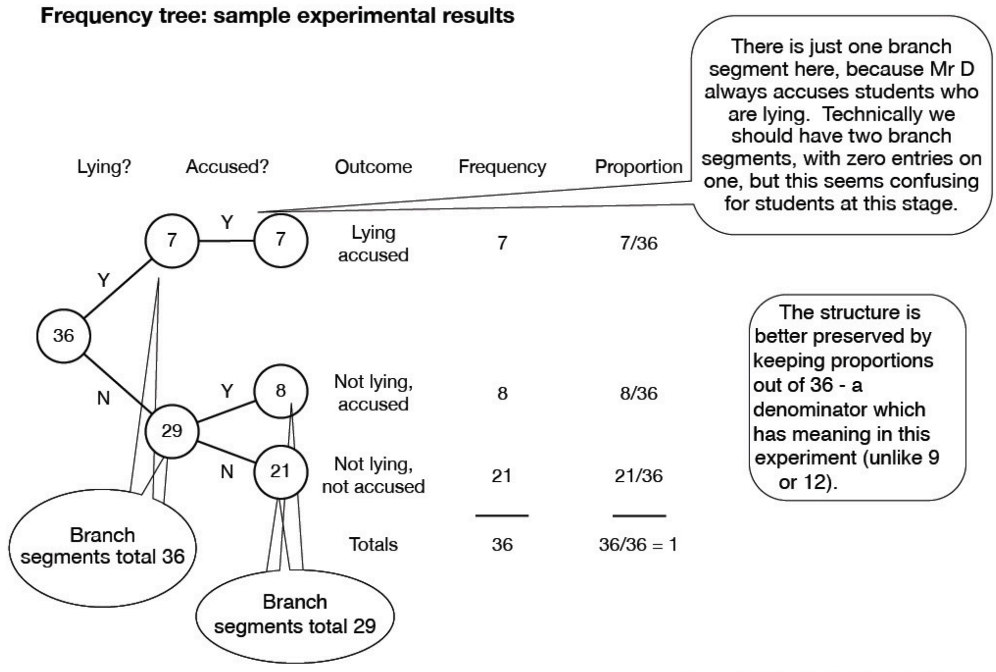


Figure 3. Empirical frequency tree of 36 independent experiments

Students should now have grasped the long-run pattern that would be expected in very large numbers of repetitions, helped by simulation software if available. They are then asked to consider what they would 'expect' to happen in 36 further experiments, as in Figure 4.

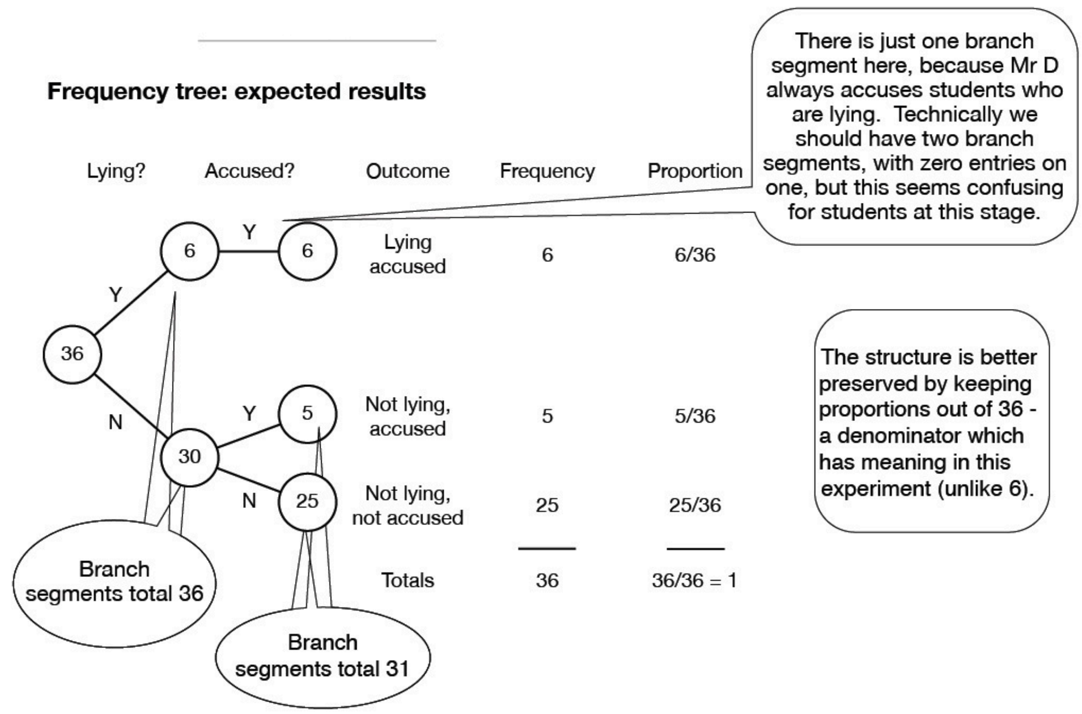


Figure 4. Expected frequency tree for a future 36 independent experiments

For more advanced students, 'Bayesian' questions can be answered by examining the expected frequency tree. For example, what proportion of accused are expected to be innocent? It is straightforward to see that, out of 11 accused, 5 will be innocent, a proportion of 5/11 or 45%. This is a difficult and unintuitive calculation, made straightforward by this representation. Gigerenzer's team has shown that doctors and other professionals can be taught to accurately carry out such Bayesian tasks using expected frequency trees (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz & Woloshin, 2007). These have been adopted in animations such as out screening tool on Understanding Uncertainty (Screening tests, n.d.).

The ambiguity of the question '*what's the chance of being falsely accused?*' can then be explored. There are 5 people in the tree being falsely accused. Do we consider these as a fraction of the whole 36, of the 11 accused, or of the 30 who were not lying? This reveals the great care needed in specifying the 'reference class', which provides the denominator, for the probability statement, and the almost universal ambiguity of questions starting '*what's the chance…*'

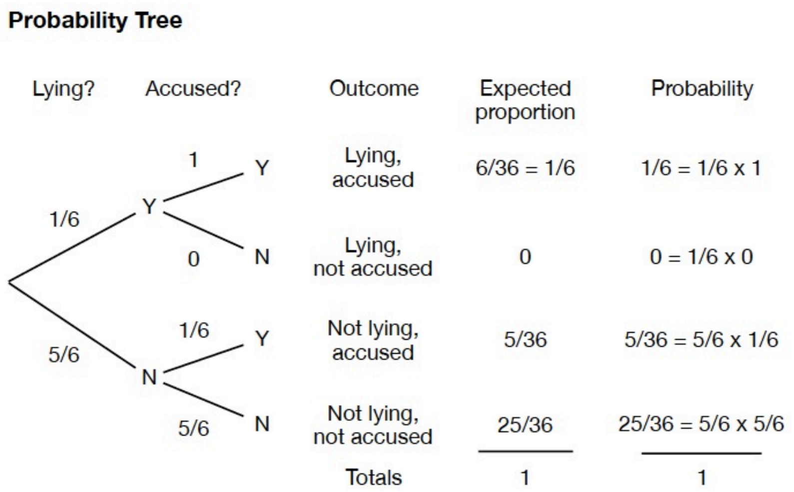


Figure 5. Probability tree obtained from expected frequency tree by examining fractions at each branch. The multiplication and addition rule of probability arise naturally.

As the final stage, the fractions at each branch can be interpreted as 'probabilities'. The multiplication rule down the branches, and the addition rule across branches, arises naturally by considering the frequencies.

Conclusions

Probability is tricky, and it is fine to think metaphorically. Recent psychology research suggests that people can be helped to solve risk problems using frequency, narrative, and multiple representations. A suggestion for teaching 'problem-solving' maths is to adopt our 'manifesto' for developing ideas through empirical frequency trees, to expected future events, and finally to probability trees. Thus probability comes at the end of the discussion, after expectation, which I personally believe that is a more fundamental concept. Further discussion and some more examples of this process can be found on the blog post (Spiegelhalter, 2014).

However, this all requires further empirical testing.

References

Breast screening benefits and harms. (2012, October 10). Retrieved November 4, 2012, from http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/screening/Benefits-and-Harms/#balance

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 1083-1092.

Galesic, M., & Garcia-Retamero, R. (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine, 170*(5), 462-468.

Gigerenzer, G. (2003). Simple tools for understanding risks: from innumeracy to insight. *BMJ*, 741-744.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53-96.

Great expectations: probability through problems. (n.d.). Retrieved from http://nrich.maths.org/probability

Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: graphical and analog tools. *Annals of the New York Academy of Sciences, 1128*(1), 18-28.

Martignon, L., & Kurz-Milcke, E. (2006). Educating children in stochastic modeling: games with stochastic urns and colored tinker-cubes. Retrieved from http://www.redeabe.org.br/ICOTS7/Proceedings/PDFs/ContributedPapers/C443.pdf

NHS breast screening. (n.d.). Retrieved April 23, 2014, from http://www.breakthrough.org.uk/about-breast-cancer/touch-look-check/nhs-breast-screening

Screening tests. (n.d.). Retrieved March 27, 2015, from http://understandinguncertainty.org/screening

Spiegelhalter, D. (2014, September 13). Using expected frequencies when teaching probability. Retrieved from http://understandinguncertainty.org/using-expected-frequencies-when-teaching-probability

Willey, J. (2012, January 12). Daily fry-up boosts cancer risk by 20 per cent. Retrieved March 27, 2015, from http://www.express.co.uk/posts/view/295296/Daily-fry-up-boosts-cancer-risk-by-20-per-cent

Probability, Justice, and the Risk of Wrongful Conviction

Jeffrey S. Rosenthal
University of Toronto, Canada

Abstract: We consider the issue of standards of proof in legal decisions from the point of view of probability. We compare ``balance of probabilities'' and ``beyond a reasonable doubt'' to the statistical use of p-values. We point out various fallacies which sometimes arise in legal reasoning. And we provide several examples of legal cases which involved probabilities, including some in which incorrect decisions were made and defendants were wrongfully convicted.

Keywords: probability, statistics, p-value, balance of probabilities, beyond a reasonable doubt, standard of proof.

Background and Context

I am a professor of statistics, and most of my work is fairly technical and mathematical (see www.probability.ca/jeff/research.html). But one day I wrote a book, *Struck by Lightning: The Curious World of Probabilities*, for the general public, which did unexpectedly well, especially in Canada. I was then interviewed by the media about such diverse topics as lottery probabilities, public opinion polls, violent crime rates, sports statistics, and more, and was even involved in probing a major lottery retailer fraud scandal involving millions of dollars and criminal convictions (see, for example, www.probability.ca/lotteryscandal). I was also invited to give talks to all sorts of different groups, from insurance brokers to financial investors, from humour therapists to gambling addiction counselors.

And then one day I was invited to speak to a prominent group of Canadian lawyers and judges. This invitation in turn led to others, and I ended up giving five different talks to five different groups of lawyers and judges (including the Irish Supreme Court and High Court justices) within a single year. This forced me to investigate the connection of probabilities and statistical analysis to the justice system, as I will now discuss.

Probability and Justice?

What is the connection of probability and statistics to justice issues? Well, both topics involve *evaluating evidence*, i.e., determining whether the available information is sufficient to draw certain conclusions. This perspective was nicely summarised by the James Bond villain Auric Goldfinger when he stated, “Once is happenstance. Twice is coincidence. The third time it’s enemy action.”

In probability and statistics, the possible conclusions might involve medical testing, or psychological analysis, or social science studies. We might try to determine if an observed difference is “statistically significant”, or if a certain probability is above some threshold, or if a corresponding p-value is less than some cutoff. In the justice system, the possible conclusions involve such questions as guilt or innocence. A court is then tasked with determining if a case has been proven “beyond a reasonable doubt”, or (for civil cases) by a “balance of probabilities” or a “preponderance of the evidence”. So what do these terms mean, and how are they related?

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 11-18

The justice term “balance of probabilities” seems closest to the world of probability and statistics. It seems to mean that a certain conclusion is more likely than not. So perhaps that means simply that the probability has to be more than 50%? Unfortunately even this connection is not so clear-cut. A classic example involves 1,000 people attending an event at which only 499 admission fares were collected. This means that a randomly-chosen attendee has probability 50.1% of not having paid admission. But surely no judge would award damages against a randomly-chosen attendee on that basis. Thus, even the simple-seeming “balance of probabilities” standard requires human judgement and cannot be interpreted in purely probabilistic terms.

The phrase “beyond a reasonable doubt” is even more challenging. It is generally agreed to mean something weaker than “certainty”, but something stronger than just “probably”. (For example, the Ireland Director of Public Prosecutions web site states: “The judge or jury has to be convinced beyond a reasonable doubt that a person is guilty. It is not enough for them to think that the accused is probably guilty.”) So does this mean the probability of guilt has to be more than 95%? more than 99%? more than 99.9%? Or that the corresponding p-value (i.e., the probability that we would have observed such evidence even if the accused were innocent) must be less than 5% or 1% or 0.1%? Once again, there is no clear standard, and the correspondence between probability/statistics and the standards of the justice system are hard to pin down.

Nevertheless, despite these challenges, it does seem that justice issues should be somewhat analysable in terms of probabilities and statistics. There are two major *risks* that need to be avoided: the risk of letting a guilty person go free, and (perhaps worse) the risk of wrongly convicting an innocent per- son. To explore these *risks* further, we next review some statistical practices, and then apply them to some specific legal cases.

How Statisticians Weigh Evidence

Consider a concrete example. Suppose your friend claims that she can distinguish Coke from Pepsi by taste alone. To test this claim, you pour Coke and Pepsi randomly into a series of glasses, and ask her to identify them. Suppose she identifies the first glass correctly (“Coke!”). This provides only slight evidence of her abilities, since she could have just gotten lucky. If she then also identifies a second glass correctly (“Pepsi!”), the evidence starts to increase. How many glasses in a row must she identify correctly before you would be convinced?

The classical statistical approach to this problem is to consider the *p- value*, i.e. the probability of observing such a result *if* your friend has no actual abilities and is just guessing randomly. So, if she guess right just once, then the p-value equals 1/2, or 50%. If she guesses right *twice* in a row, the p-value becomes $(1/2) \times (1/2) = 25\%$, where we multiply because (assuming random guessing) the different guesses are *independent*. If she guesses right *five* times in a row, the p-value equals $(1/2) \times (1/2) \times (1/2) \times (1/2) \times (1/2) \approx 3.1\%$.

Clearly, smaller and smaller p-values start to suggest that your friend’s guessing wasn’t just luck, but rather showed some true ability. But how small should the p-value be, to actually “prove” some conclusion? The usual standard, used throughout the medical and social sciences, is that a result is “statistically significant” if the p-value less than 5%, i.e. less than one chance in 20. Indeed, each time you take a medical drug, you are almost certainly consuming something which has been approved for treatment based on some study with some p-value which is less than 5%.

By this standard, if your friend guesses Coke versus Pepsi correctly twice in a row then this proves nothing, while if she guesses correctly five times in a row then this provides statistically significant evidence of her abilities. And, exactly the same reasoning applies to a new cure for a disease with a 50% fatality rate, which manages to save five patients in a row.

So far so good. But such statistical reasoning is not infallible. Sometimes the "evidence" is misleading due to incorrect reporting or biased sampling or incomplete recording. And even if the evidence is correct, sometimes the calculations or the conclusions are not.

One issue that arises is the *When To Multiply* question. That is, when is it appropriate for probabilities be *multiplied*? For example, if the probability of heads on one coin flip is 1/2, then the probability of heads on two coin flips is (1/2) × (1/2) = 1/4, because the coin flips are independent so multiplica- tion is valid. And in the above Coke/Pepsi example, if your friend is guessing without ability and the drinks are poured randomly, then each guess is independent of the next, so multiplication is again valid. But not always! For example, 49.2% of Americans are male, and 64% of Americans watch NFL football (according to a recent survey). So, does this mean the percentage of Americans who are male and watch NFL football is 49.2% × 64% = 31.5%? No, it's actually 49.2% × 73% = 35.9%, since 73% of American males (and only 55% of American females) watch NFL football. That is, gender and football are not independent, so the multiplication is invalid and leads to too small a probability. I will consider this further below.

In addition, it is important to *interpret* p-values correctly. If your friend guesses Coke/Pepsi correctly five times in a row, then the p-value is 3.1%. This means that *if* your friend was guessing randomly, then the probability they would perform so well is 3.1%. This does *not* mean that the probability your friend was guessing randomly is only 3.1%. These two different probabilities are often conflated, which is sometimes called the *Prosecutor's Fallacy*. In fact, the probability your friend was guessing randomly cannot be determined based on the experiment alone – it also depends on what you know or assume about your friend, and many other factors. In any case, it is *not* the same as the p-value, and indeed it could be very different.

Of even greater concern is *multiple testing*, or what I call the *Out Of How Many* principle. For example, in my book, *Struck by Lightning*, I tell the true story of running into my father's cousin at Disney World, an event which seemed like one chance in hundreds of million. However, when you consider all the strangers I saw on that trip to Disney World, as well as all the people I *would* have been surprised to run into, it turns out that probability of running into *someone* surprising during a two-day trip to a crowded location is actually more like 0.5%, i.e. not so surprising after all.

In the context of the Coke/Pepsi experiment, if your friend keeps trying to guess all afternoon, and *eventually* guesses correctly five times in a row, then this proves nothing because they had so many chances to achieve this result.

Similarly, someone winning a lottery jackpot isn't necessarily suspicious even though they have defied odds of one in tens of millions, because of all the *other* people who bought a lottery ticket and *could* have won instead. The same reasoning applies to all sorts of coincidences and suspicious – or surprising – seeming occurrences. That is, an apparently small p-value should always be treated with caution when other equally-surprising events were also possible.

The Prosecutor's Fallacy, and the When To Multiply question, and the Out Of How Many principle, all have important applications to legal cases – as we now discuss.

A Legal Case: Sally Clark

Sally Clark was a solicitor in Cheshire, England. She had two sons, each of whom died in infancy with no apparent cause. The first death had been ruled a "cot death", i.e., a case of Sudden Infant Death Syndrome (SIDS) in which babies sometimes suffocate without apparent explanation. But when the second infant also died, suspicions were raised, and Sally Clark found herself charged with double murder.

The prosecution case rested on probabilities. At her 1999 trial, the pae- diatrician Sir Roy Meadow testified that "the odds against two cot deaths in the same family are 73 million to one". Clark was convicted on this basis, put in prison, and vilified in the press. She even had a third son temporarily removed from her custody. But was her conviction justified?

One issue is how the figure "73 million to one" should be interpreted. To the casual observer – or to the media, or even to a judge or juror – this might seem to be the probability that Clark is innocent. But the figure is actually a p-value, i.e. the probability that a law-abiding parent would have two infants die without apparent explanation, which is a rather different thing. Confusing the two is a classic example of the Prosecutor's Fallacy!

A second issue is whether the figure "73 million to one" was computed correctly. Meadow obtained this figure by first saying that the probability of *one* child dying of SIDS was one chance in 8,543, and then saying that for *two* children we have to multiply to get a figure of $(1/8,543) \times (1/8,543) = 1/72,982,849 \approx 1/73,000,000$. However, this multiplication was not valid. Indeed, SIDS tends to run in families, so once a family has had one SIDS case, the second one is more likely – just like for male football watchers.

Furthermore, even the figure 1/8,543 for each individual SIDS death was not valid. The overall probability of SIDS in the U.K. had been estimated as 1/1,303. Meadow obtained 1/8,543 by "adjusting" for family circumstances that lower the SIDS probability (e.g. no smokers, someone employed, mother over 26 years old). But he neglected other factors which *raise* the probability (e.g. that SIDS is twice as likely for boys as for girls). So, for all of these reasons, the correct probability of two SIDS deaths in the same family was surely higher than the "one in 73 million" figure used in court.

But most important of all is the Out Of How Many principle. After all, there are tens of millions of families in U.K. alone. So, the fact that *one* of them had two SIDS deaths is not so surprising – much like the probability that *someone* wins the lottery jackpot. To convict solely on the basis of probabilities, the chances would have to be so low that we would *never* expect to see such an occurrence even once in any family in the U.K. or perhaps the entire world. By that standard, even a tiny p-value like "one in 73 million" is, well, not tiny enough. (By contrast, if Sally Clark was *already* under suspicion for some *other* reason, then a small p-value could well be convincing, since then the Out Of How Many principle might not apply.)

Sally Clark was convicted of double-homicide in November 1999, purely on the basis of Meadow's probability calculation. However, statisticians soon noticed the flaws in the case. The venerable Royal Statistical Society noted (see: www.rss.org.uk/uploadedfiles/documentlibrary744.pdf) that Meadow's approach was "statistically invalid", and declared that "The case of R v. Sally Clark is one example of a medical expert witness making a serious statistical error, one which may have had a profound effect on the outcome of the case." For these and other reasons, Clark was ultimately acquitted on her second appeal, after more than three years in jail. But she never recovered psychologically, and died of alcohol poisoning four years later.

Meanwhile, the U.K. General Medical Council ruled that Meadow's evidence was "misleading and incorrect", and that he was guilty of "serious professional misconduct". He was effectively barred from any future court work. Furthermore, the prosecution pathologist Alan Williams was found to have not reported evidence about an *infection* in the second son (which may have suggest death by natural causes). The GMC found him, too, guilty of serious professional misconduct. As a result, several other similar convictions were also overturned on appeal. A valuable lesson had been learned.

An Earlier Case: Malcolm Collins

On June 18, 1964, in Los Angeles, an elderly lady was pushed down in an alley, and her purse was stolen. Witnesses said: a young Caucasian woman, with a dark blond ponytail, ran away with the purse, into a yellow car, which was driven by a Black man, who had a beard and moustache. Four days later, Malcolm and Janet Collins were arrested, primarily because they fit these same characteristics (at least mostly – Janet's hair was apparently light blond rather than dark blond).

At trial, the prosecutor called "a mathematics instructor at a nearby state college" (whose identity no one seems to know). The prosecutor told the mathematics instructor to assume certain "conservative" probabilities:

•Black man with a beard: 1 out of 10

•Man with moustache: 1 out of 4

•White woman with blond hair: 1 out of 3

•Woman with a ponytail: 1 out of 10

•Interracial couple in car: 1 out of 1,000

•Yellow car: 1 out of 10

The mathematics instructor then computed the probability that a random couple would satisfy all of these criteria, by – you guessed it – multiplying:

$(1/10)\times(1/4)\times(1/3)\times(1/10)\times(1/1000)\times(1/10) = 1/12{,}000{,}000$

It was thus asserted that there was just one chance in 12 million that a couple would have these same characteristics if they were not guilty. Malcolm Collins was convicted at trial, primarily based on this "one in 12 million" probability.

Was this probability calculation valid? Surely not. For one thing, those individual probabilities were just *assumed*, without evidence. Furthermore, the multiplication was again invalid: for example, most men who have beards also have moustaches, so (like the male football watchers) these factors are surely not independent. (And, if you have a Black man and a White woman, then *of course* you have an interracial couple! Perhaps the prosecutor may have meant that one in 10 Black men have beards, not that one man in 10 is Black with a beard, but the interpretation is rather difficult to sort out.) So, the asserted probability of "one in 12 million" is highly questionable.

Even more important, once again, is the Out Of How Many principle. Los Angeles County in 1964 had a population of 6,537,000, and thus approximately one million couples (which could termed the "suspect population"). So, even with odds as small as one in 12 million, the probability of there being two such couples is actually fairly large. To convict Malcolm Collins on the basis of probabilities alone seems inappropriate in this case.

The case of Malcolm Collins eventually made its way to the Supreme Court of California. Their 1968 judgment (found here: scholar.google.com/scholarcase?case=2393563144534950884) began, "We deal here with the novel question whether evidence of mathematical probability has been properly introduced and used". They rightly observed that "the testimony as to mathematical probability infected the case with fatal error". However, they then further insisted that the trial's probability calculations had "distorted the jury's traditional role of determining guilt or innocence according to long-settled rules", concluding that "Mathematics, a veritable sorcerer in our computerized society[1], while assisting the trier of fact in the search for truth, must not cast a spell over him. We conclude that on the record before us defendant should not have had his guilt determined by the odds". They overturned the conviction on that basis. Now, I am glad that the conviction was overturned, due to the flaws in the probabilistic reasoning. But I wish they hadn't implied that guilt should *never* be determined by the odds – I disagree and think that is going too far.

Another Case: Lucia de Berk

Lucia de Berk was a nurse who worked on three different hospital wards in The Hague, Netherlands. She was arrested after it was discovered that she was on duty for 14 of 27 "incidents" (i.e. patient deaths or near-deaths) in her three wards (51.9%), despite working just 203 of the 2,694 shifts in her three wards (7.5%). At her trial, the prosecution asserted that there was just one chance in 342 million that such an imbalance would occur by chance alone. de Berk was convicted of multiple murders and attempted murders in March 2003, primarily on the basis of this "1 in 342 million" probability. Was her conviction justified?

A first question is whether the evidence (i.e. facts) were accurate. There was some controversy about whether all of these incidents had actually taken place *during* de Berk's shifts, as opposed to just before or just afterwards. Furthermore, the definition of "near-death" might have been adjusted *post hoc* to include more incidents during de Berk's shifts. Related to this, de Berk may have been assigned to extra elderly/terminal patients due to her experience as a nurse, which may have provided an alternative explanation of any excess number of incidents. These issues were all debated vigorously following her conviction.

In addition to the above, many of our previous concerns apply. Once again, the Prosecutor's Fallacy must be avoided: the probability that de Berk is guilty *given* the observed facts is quite different from the probability that the observed facts *would* have arisen if she were innocent. Even more important is the Out Of How Many principle. The prosecution statistician, Henk Elffers, had tried to account for this by multiplying by 27 (the number of nurses in one of the hospitals), but arguably he should really have multiplied by the number of nurses in the entire Netherlands or even the whole world.

After de Berk's conviction, various statisticians objected. In particular, four Dutch statisticians alluded to the Out Of How Many principle by saying "the data . . . is used twice: first to identify the suspect, and after that again in the computations of Elffers' probabilities" (Meester, Collins, Gill & Lambalgen, 2006). They made numerous "adjustments", and eventually increased the p-value from "1 in 342 million" to 0.022 (i.e. 1 chance in 45), a p-value which is surely too large for conviction.

de Berk's convictions were upheld on first appeal in 2004. However, enough doubts had been raised about the probability calculations that the conviction was upheld primarily on *other* grounds, notably elevated digoxin levels in some of the corpses (which could be evidence of poisoning).

---

[1] If they thought society was "computerized" in 1968, what would they think today?

However, the hypothesis of digoxin poisoning was disproven by 2007, leading to the case being reopened in 2008, and a not guilty verdict being delivered on second appeal in 2010. Lucia de Berk is now a free woman.

Of course, none of this precludes the possibility that Lucia de Berk might have been guilty. For example, she may have killed some terminal patients out of *mercy*, to relieve their suffering in their final days. Indeed, on the day of one of her elderly patient's death, de Berk wrote in her diary that she had "given in to her compulsion" (though she later claimed she was referring to her compulsion to read Tarot cards). While this fact was introduced at her trial, her conviction was based primarily on the statistical evidence. And, as we have seen, the statistical evidence wasn't sufficiently convincing.

Discussion

I have presented three different legal cases where people were convicted of serious crimes primarily on the basis of faulty probability calculations. It may be tempting for some people to conclude from this – as the Supreme Court of California perhaps did in 1968 – that probabilities should never be used to convict anyone of anything.

I think that this is going too far, and that statistical analysis *can* sometimes help to achieve justice after all, provided that it is used with caution. One example of this is the lottery retailer scandal mentioned earlier. In that case, I was able to determine that lottery retailer ticket sellers had won more major lottery prizes than could be reasonably explained by chance alone. This conclusion became a huge news story in Canada, and led to millions of dollars in lottery repayments, and several criminal convictions for fraud (see www.probability.ca/lotteryscandal). This illustrates how careful statistical calculations which take into account the factors mentioned above can identify criminal activity and achieve justice.

An interesting related story is that of Waneta & Tim Hoyt. They had five babies in New York State during 1965 – 1971, *all* of whom died in infancy (at 3, 28, 1.5, 2.5, and 2.5 months old, respectively). The deaths were all identified as SIDS, and indeed a pediatrician used them to publish a scholarly article about SIDS' strong genetic linkage (Steinschneider, 1972). Apparently no foul play was suspected, and in fact the Hoyts were later allowed to adopt a son (who survived to adulthood) in 1977. Years later, in 1985, some prosecutors and pathologists became suspicious, and investigated. Eventually, Waneta Hoyt confessed to suffocating all five of the children, to stop them from crying. She later "recanted" her confession, but was nevertheless convicted in 1985 of five murders; she died in prison in 1998 at the age of 52. It would appear, at least in hindsight, that her murderous ways should have been detected much sooner – but instead the genetic linkage was believed so strongly that even five deaths were not considered suspicious.

As these examples illustrate, it is difficult to decide when probabilistic evidence is sufficient to justify a criminal conviction. The calculation and interpretation of p-values is often challenging. Overly aggressive or simplistic calculations run the risk of convicting innocent people, while overly cautious analyses run the risk of setting guilty parties free. Alternatively, it is possible to take a Bayesian approach to this question (see, for example, Meester et al., 2006, section 6), but that requires specifying prior probabilities which is itself problematic and subjective.

Nevertheless, I do believe that probabilities can and should be used in criminal trials (among other places). Such probabilities must be carefully computed, accounting for such issues as the accuracy of the data, the When To Multiply question, the Prosecutor's Fallacy, and (perhaps most important of all) the Out Of How Many principle. If all of these factors are carefully taken into account, then probabilities can indeed be used to draw conclusions and avoid risks, even about criminal activity.

References

Meester, R., Collins, M., Gill, R., & van Lambalgen, M. (2006). On the (ab)use of statistics in the legal case against the nurse Lucia de B. Law, *Probability and Risk, 5*, 233–250.

Steinschneider, A. (1972). Prolonged apnea and the sudden infant death syndrome: clinical and laboratory observations. *Pediatrics, 50*(4), 646–654.

Risk: Mathematical *and* Otherwise

John Adams
University College London, England

Abstract: What role might mathematicians have to play in the management of risk? The idea of turning a risk, a *possibility* of loss or injury, into a "calculated" risk, a quantified *probability* of loss or injury, is one that has obvious appeal not just to statisticians and mathematicians – but to large numbers of others who would like to know the probability of failure before pursuing some intended course of action. Conclusion: even when risks can be calculated with great precision, they can only be used to inform judgment, but not substitute for it. And it matters *who* is making the judgment.

Keywords: risk compensation, virtual risk, probability, risk amplification.

"It was a calculated risk, and we forgot to carry the one."

[Thanks to Mark Anderson for permission to reproduce]

In 1999 NASA's Mars Climate Orbiter burned and crashed because no one had thought to check whether force, expressed in pounds, had been converted to force expressed in Newtons (Grossman, 2010). Failure to carry the one, or convert pounds to Newtons, are examples of only one of the risks encountered in attempting to apply mathematics to the management of risk.

Assuming they can remember to carry the one, what role might mathematicians have to play in the management of risk? The idea of turning a risk, a *possibility* of loss or injury, into a "calculated" risk, a quantified *probability* of loss or injury, is one that has obvious appeal not just to statisticians and mathematicians – but to large numbers of others who would like to know the probability of failure before pursuing some intended course of action.

"Risk" (almost a billion Google hits) has become a booming business. "Risk management" yields over 80 million hits, and "chief risk officer" (of interest to those looking for employment in this field) returns half a million. Governments are keen on risk management: Turnbull, Basel, Sarbanes-Oxley are names associated with guidance, accords or legislative acts intended to ensure that financial risks are managed effectively. Most big banks now have extraordinarily highly paid chief risk officers (CROs) to ensure compliance with their requirements – in 2011 the CRO at Bank of America was paid $11.4 million (Bloomberg News, 2011). Other large, non-financial, enterprises such as General Motors and Ford, Shell and BP, Delta Airlines, Toyota, also have senior executives bearing the CRO title.

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 19-30

The financial meltdown of 2007/2008 gave a huge boost to the risk management industry. It has now declared itself a profession and it is growing at an impressive rate: GARP, the Global Association of Risk Professionals grew more than three-fold from 55,000 members pre-crash in 2006 to more than 175,000 by 2011.

Types of Risk

The growing army of risk managers seeks to manage an extraordinary range of different risks. Here is a short starter list: financial risk (credit risk, market risk, liquidity risk, value at risk …), legal risk, reputation risk, medical risk, strategic risk, policy risks, inflation risk, recession risk, terrorism risk, sanctions risk, climate risk, radiation risk, extreme weather risk, road accident risk, etc., etc.

The list could go on almost without end. Any threat of nature or any human activity, physical or intellectual, leading to an uncertain outcome can serve as a descriptor of a type of risk.

A further, less open-ended, set of categories can be helpful in an attempt to illuminate the challenges facing risk managers seeking to reduce risks to calculable probabilities. Figure 1 presents a risk typology that is germane to most discussions of a wide variety of risks and their management.

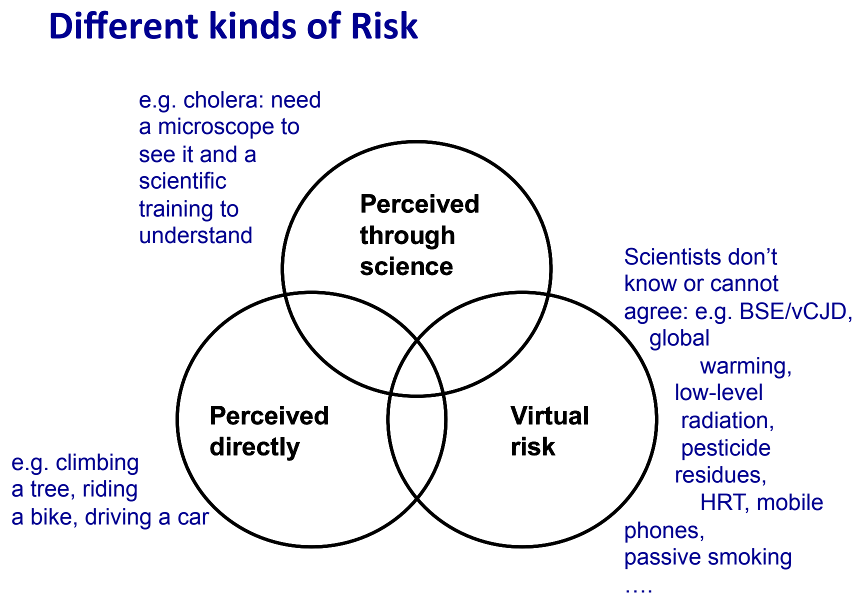


Figure 1. Different kinds of risk

The Venn diagram in Figure 1 suggests that the typology can be useful to distinguish three different, but not mutually exclusive, types of risk. One need sample only a tiny fraction of the 100s of millions of Google "risk" hits to discover unnecessary and often acrimonious arguments caused by people using the same word to refer to different things and shouting past each other. The typology offered in Figure 1 can help to dispose of some unnecessary arguments and, perhaps, civilize others.

Risks in the *perceived directly* circle are managed using *judgement*. We do not undertake a formal, probabilistic risk assessment before crossing the road; some combination of instinct, intuition and experience usually sees us safely to the other side. The consequences of failing to carry the one, or convert pounds to Newtons are, like road accidents, usually the result of carelessness: a failure pay attention to directly perceptible hazards.

The second, the *risk-perceived-through-science* circle, dominates the risk management literature. This is the circle within which most of the risk professionals ply their trade. It is the mathematical circle. In this circle we find books, reports and articles with verifiable numbers, cause-and-effect reasoning, probability and inference. This is the domain of, amongst others, biologists with microscopes searching for microbial pathogens and astronomers with telescopes plotting the courses of incoming asteroids. This circle contains contributions from the whole range of science, technology and the social sciences – from physics and chemistry to epidemiology and criminology. But the central science is statistics – the discipline that has probability at its core. The future can be imagined with the help of statisticians, but only *if one is happy to assume* that the historic relationships embodied in their models will persist unchanged into an uncertain future.

The circle labelled *virtual risk* contains contested hypotheses, ignorance, uncertainty and unknown unknowns. If an issue cannot be settled by science and numbers, we rely, as with directly perceptible risks, on *judgement*. Some find this enormously liberating; interested parties are freed to argue from their beliefs, prejudices or superstitions. It is in this circle that we find the longest-running and most acrimonious arguments. Virtual risks may or may not be real, but beliefs about them have real consequences. Global warming has been placed in this circle because the (potentially catastrophic?) warming of which some warn, and which others dispute, is the product of models that grossly simplify extremely complex systems, but lead some to propose policies that would, if pursued, dramatically alter the life-styles of billions.

Risk on the Road: Numbers, and Arguments about Numbers

We can find all three of these risk types contending on the road. In order to contain the discussion in this essay within reasonable bounds I will focus mainly on examples from the realm of road safety. Road safety is an issue that comes with a large number of numbers attached. And they settle few arguments.

People living alongside roads with high volumes of fast traffic often complain, relying on their *direct perceptions*, that their roads are dangerous, and campaign for measures that will reduce the volume and slow the speed of the traffic outside their front doors. Their campaigns sometimes bring them into contact with the highway engineers responsible for their roads. The engineers are likely to confront them with numbers from the *mathematical* circle of Figure 1. Their road accident hot-spot maps show that the roads complained of are safe, with no, or very few, accidents. But the people living alongside the road are unpersuaded by the numbers on the engineers' maps. They can *see* that their roads are dangerous.

The road environment also throws up numerous problems that can be consigned to the "virtual" circle of Figure 1 – issues about which people cannot agree or confess ignorance. At what age is it safe to allow your children to get to school on their own, or cross a busy road? How will driverless cars interact with pedestrians and cyclists? Should cyclists be compelled to wear helmets, or motorists seat belts[1]? All these are current on-going debates that spring to mind.

Crossing the road

I offer Figure 2 as a simple model of what goes on in my head when I am crossing the road. I call it the risk thermostat. The thermostat is set in the top left-hand corner. The setting of risk thermostats can vary enormously – from that of a timid and cautious little old lady named Prudence to that of a wild and reckless Hell's Angel. But everyone has some propensity to take risks; a zero risk life is not possible.

[1] Readers who thought that this debate had been settled are referred to "Britain's seat belt law should be repealed" (http://john-adams.co.uk/wp-content/uploads/2008/08/seat-belts-for-significance-2.pdf). It is a debate that goes back to 1982 (see, for example, http://john-adams.co.uk/wp-content/uploads/2006/SAE%20seatbelts.pdf) and more recently (see, for example, http://www.john-adams.co.uk/?s=seat+belts)

A *propensity* to take risks leads to risk taking behaviour that leads, by definition, to *accidents*. To take a risk is to do something that carries with it a probability of an adverse outcome. Through having accidents, and surviving them and learning from them, or seeing them on television, or being warned long ago by mother, I have acquired a *perception* of the risks associated with crossing roads. The model proposes that when my perception of a risk and my propensity to take it are out of balance I change my behaviour to restore the balance. Why do I cross the road? To get to the *reward* on other side; and the magnitude of that reward will influence the setting of my thermostat. The change in behaviour in response to changes in the perception of risks described by Figure 2 is commonly known as *risk compensation.*

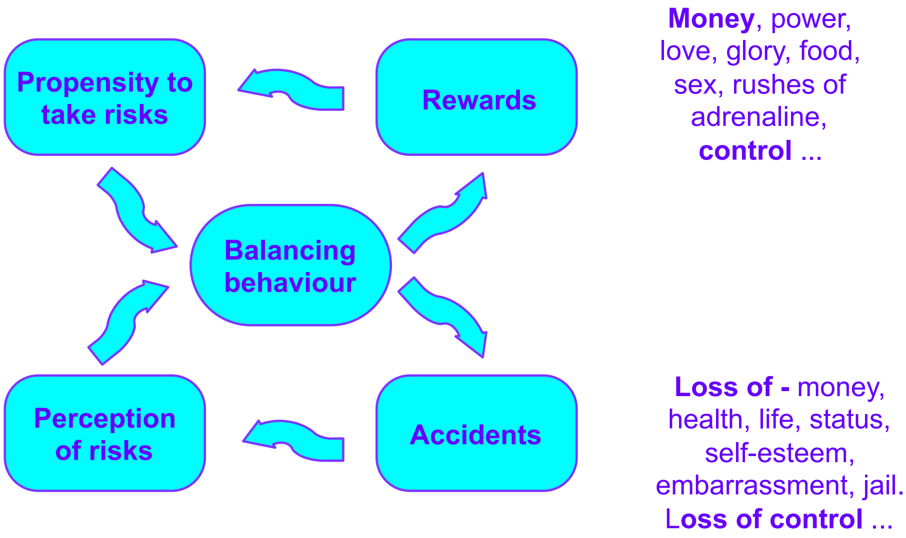


Figure 2. The risk thermostat

I used to describe the process illustrated by Figure 2 as cost-benefit analysis without the £ or $ signs, but nothing, it appears, is beyond the determined economist's ability to be rendered as a sum of money. Spending on drugs and other medical procedures is commonly justified in terms of the Qalys (quality adjusted life years) that they would yield – the UK value of a Qaly currently ranges from £10,000 to £70,000 (Donaldson et al., 2011). A significant benefit claimed for new road schemes in Britain is the value of the lives that they would save – with each life currently valued at £1,249,890 (2006 value, routinely adjusted for inflation: http://webarchive.nationalarchives.gov.uk/20100304070241/http://www.dft.gov.uk/pgr/economics/software/coba11usermanual/part2thevalofcostsandb3154.pdf). And the Stern Report, an influential contribution to the climate change debate in Britain has sparked a debate framed in terms of the monetary costs and benefits, and their discount rates, likely to be incurred or enjoyed many generations into the future (see, for example, http://en.wikipedia.org/wiki/Stern_Review#The_costs_of_mitigation).

What kills you matters

But it appears that in the eyes of many non-economists, some pounds or dollars are more equal than others. In listing some of the contents of the *Rewards* and *Accidents* boxes in Figure 2 *control* and *loss of control* have been highlighted. Figure 3 sets out the significance of this factor.

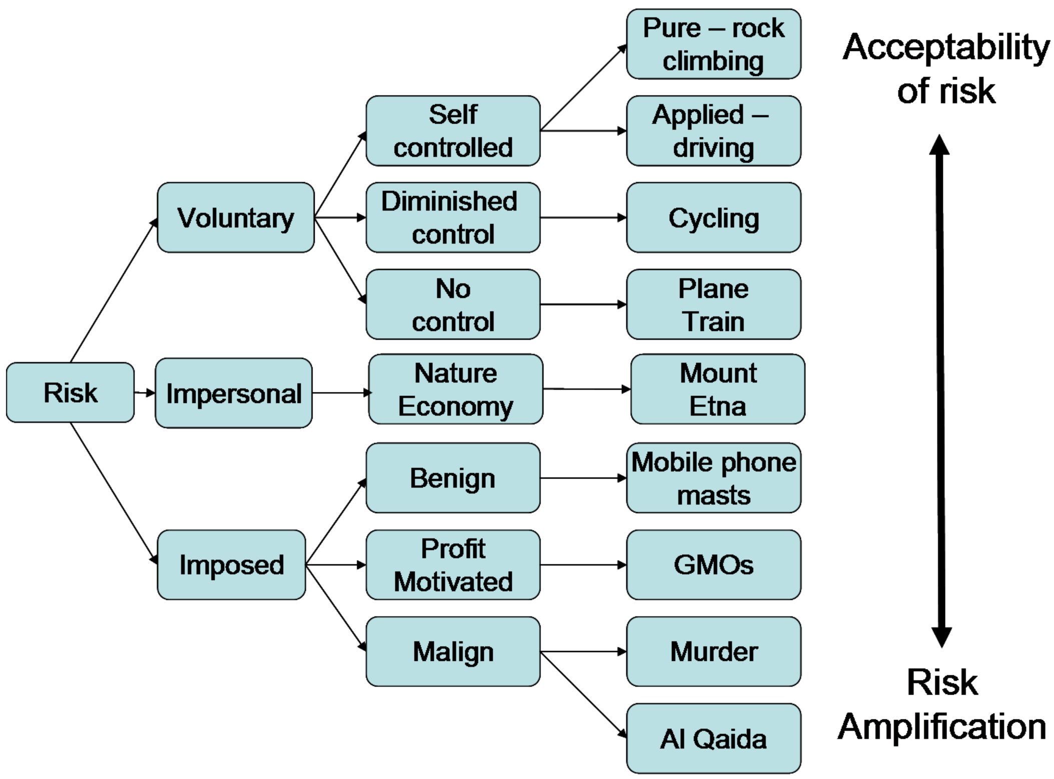


Figure 3. What kills you matters

Acceptance of a given actuarial level of risk varies widely with the perceived level of control an individual can exercise over it and, in the case of imposed risks, with the perceived motives of the imposer.

With 'pure' voluntary risks, the risk itself, with its associated challenge and rush of adrenaline, is the reward. Most climbers on Mount Everest and K2 know that it is dangerous and willingly take the risk. Similarly thrill-seeking young men driving recklessly are aware that what they are doing is dangerous; that is the point.

With a voluntary, self-controlled, applied risk, such as driving, the reward is getting expeditiously from A to B. But the sense of control that drivers have over their fates appears to encourage a high level of tolerance of the risks involved.

Cycling from A to B (I write as a London cyclist) is done with a diminished sense of control over one's fate. This sense is supported by statistics that show that per kilometre travelled a cyclist is much more likely to die than someone in a car. This is a good example of the importance of distinguishing between relative and absolute risk. Although much greater, the absolute risk of cycling is still small – 1 fatality in 25 million kilometres cycled; not even Lance Armstrong can begin to cover that distance in a lifetime of cycling. And numerous studies have demonstrated that the extra relative risk is more than offset by the health benefits of regular cycling; regular cyclists live longer.

While people may voluntarily board planes, buses and trains, the popular reaction to crashes in which passengers are passive victims, suggests that the public demand a higher standard of safety in circumstances in which people voluntarily hand over control of their safety to pilots, or bus or train drivers.

Risks imposed by nature – such as those endured by people living on the San Andreas Fault or the slopes of Mount Etna – or by impersonal economic forces – such as the vicissitudes of the global economy – are placed in the middle of the scale. Reactions vary widely. Such risks are usually seen as motiveless and are responded to fatalistically – unless or until the risk can be connected to base human motives. The damage caused by Hurricane Katrina to New Orleans is now attributed more to wilful bureaucratic neglect than to nature. And the search for the causes of the economic devastation attributed to the 'credit crunch' is now focusing on the enormous bonuses paid to the bankers who profited from the subprime debacle.

Risks imposed by one's fellow humans are less tolerated. Consider mobile phones. The risk associated with the handsets is either non-existent or very small. The risk associated with the base stations, measured by radiation dose, unless one is up the mast with an ear to the transmitter, is orders of magnitude less. Yet all around the world billions of people are queuing up to take the voluntary risk, and almost all the opposition is focused on the base stations, which are seen by objectors as impositions. Because the radiation dose received from the handset increases with distance from the base station, to the extent that campaigns against the base stations are successful, they will increase the distance from the base station to the average handset, and thus the radiation dose. The base station risk, if it exists, might be labelled a benignly imposed risk; no one supposes that the phone company wishes to harm all those in the neighbourhood. And the extent to which traffic is seen as an imposed risk varies widely; parents of young children and cyclists are much more likely to feel it as an imposition than drivers of SUVs and big cars.

Even less tolerated are risks whose imposers are perceived to be motivated by profit or greed. In Europe, big biotech companies such as Monsanto are routinely denounced by environmentalist opponents for being more concerned with profit than the welfare of the environment or the consumers of its products. Manufacturers of high-performance cars are assigned by some road-safety campaigners to the same category, their arguments sometimes adding damage to the environment to the danger posed to vulnerable road users.

Less tolerated still are malignly imposed risks – crimes ranging from mugging to rape and murder. In most countries the number of deaths on the road far exceeds the numbers of murders, but far more people are sent to jail for murder than for causing death by dangerous driving. In the United States in 2012 14,827 people were murdered – a statistic that evoked far more popular concern than the 33,561 killed on the road – but far less concern than that inspired by the zero killed by terrorists.

Which brings us to Al Qaida, Isis and their associates. How do we account for the massive scale, world-wide, of the outpourings of grief and anger attaching to its victims, whose numbers are dwarfed by victims of other causes of violent death? In London 52 people were killed by terrorist bombs on 7 July 2005, about six days worth of death on the road in the whole country. But thousands of people do not gather in London's Trafalgar Square every Sunday to mark, with a three-minute silence, their grief for the previous week's road accident victims.

The dangers that can be tracked to the malign intent of terrorists are amplified by governments who see them as a threat to their ability to govern – to their ability to control events. To justify forms of surveillance and restrictions on liberty previously associated with tyrannies, 'democratic' governments now characterize any risk to life posed by terrorists as a threat to *Our Way of Life*.

Moving from the top to bottom of Figure 3 we encounter a phenomenon known as risk amplification. *The numbers almost don't matter.* Figures 2 and 3 can also help to explain the discrepancy referred to above between the judgments of local residents and the mathematically based judgement of the highway engineer about the safety of a road. If the residents perceive their road to be dangerous they will modify their behaviour. Old people will be afraid to cross it. Children will be forbidden to cross it. And fit adults will cross it quickly and carefully. Their good accident record is often purchased at the cost of community severance. People on one side of the road tend no longer to know their neighbours on the other side. The numbers on the highway engineer's accident map measure not safety, but risk aversion. And those living on the road will tend to see the danger as an imposed risk, amplifying their perception of the risk.

Some more numbers from Great Britain

Staying with risk on the road, Figure 4 describes the rise and fall of road accident deaths in Great Britain between 1950 and 2012. I use Great Britain as an example because it is the country with which I am most familiar, but most highly motorised countries display similar patterns over this period. GB road accident fatalities reached a post-war peak of 7985 in 1966 before falling to 1713 in 2013 – the lowest since records began.

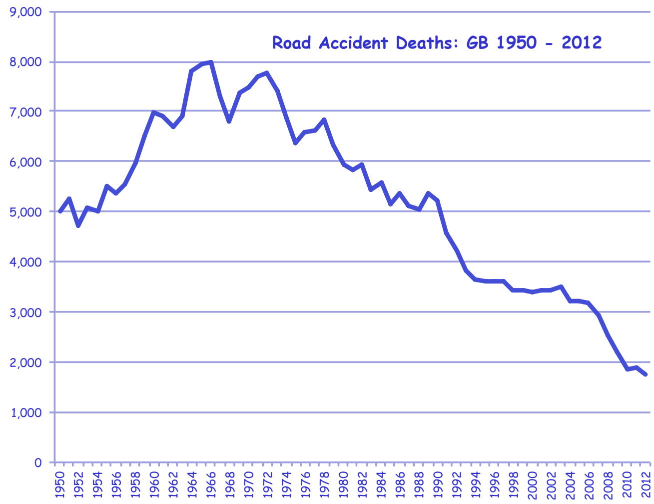


Figure 4. Road accident deaths: Great Britain (1950-2012)

How might the numbers represented by this graph be explained? Over the period traffic grew considerably; cars improved in terms of crash protection and better brakes; numerous laws were passed to curb speed, compel the use of seat belts and helmets, and ban drinking and driving and the use of mobile phones. And highway engineers lengthened sight-lines, installed central barriers on freeways and pedestrian barriers in cities, and removed roadside obstacles such as trees.

So, was it the engineers with their improved brakes and crash protection? Was it the road builders with their safer roads? Was it the legislators and legislation enforcers? Who deserves the credit for this extraordinary reduction in road accident fatalities?

Figure 4 transforms Figure 3 in a way that sheds some light on the possible risk reduction effect of all these measures. It represents fatalities per unit of exposure – i.e. per billion vehicle kilometres of traffic. It shows a 96% decrease over the 62-year period displayed on the graph,

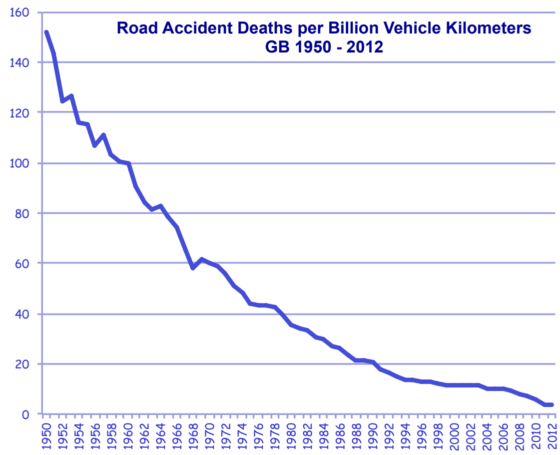


Figure 5. Road accident deaths per billion vehicle kilometers : Great Britain (1950-2012)

And Figure 5, with the vertical axis logged, transforms the graph again in a way that poses challenging questions to those who would claim credit for any of the risk reduction measures listed above.

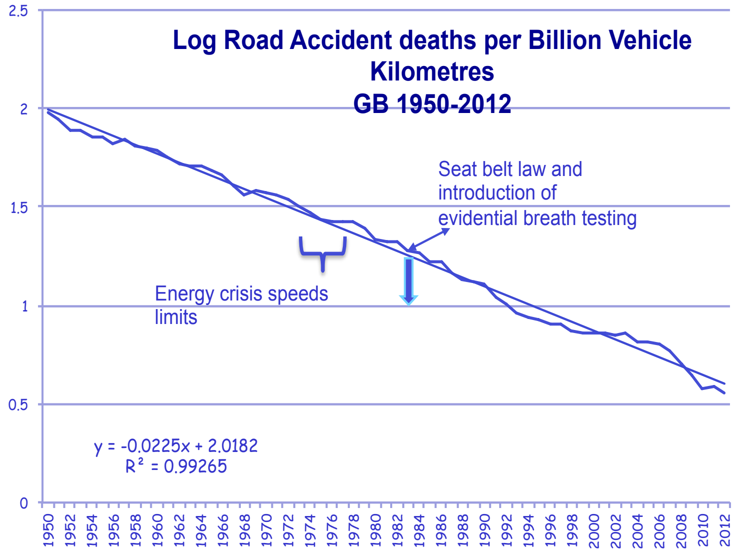


Figure 6. Log road accident deaths per billion vehicle kilometres: Great Britain (1950-2012)

The slope of the straight line indicates that, over this period, fatalities per kilometre fell by an average of 5.3% per year. Over most of the period before 1966 traffic increased faster that 5.3% per year and afterwards more slowly with the result that over the whole of the period fatalities per kilometre declined.

But it is extraordinarily difficult to spot the contributions of the vehicle engineers and manufacturers, the legislators and the road builders referred to above. Over this period the interventions whose promoters promised would have the largest and most immediate effects were the energy crisis speed limits and, in 1983, the seat belt law and the introduction, in the same year, of evidential breath-testing. Both promised instant large downward steps on the graph in Figures 5 and 6 and both are very difficult to see. The largest single step down over the whole period was in 1991 when nothing significant happened on the road safety front – except for the most severe economic recession since the war.

In Britain, over this period, the largest road safety claims, by a wide margin, have been made by the seat-belt campaigners. In 2008, the 25th anniversary of the seat-belt law, the Department for Transport, the Parliamentary Advisory Council on Transport Safety and the Royal Society for the Prevention of Accidents all published press releases claiming credit for their contribution to the creation of a law that had saved 60,000 lives over the previous 25 years.

Here we have an opportunity for mathematicians

The claims are outrageous nonsense. British mathematical enthusiasts were asleep at the switch. The downward arrow on Figure 6 illustrates the magnitude of the sharp downward step that should have occurred on the graph if the claims (averaging 2400 lives a year over 25 years) were true.

Legislators and engineers have, for many years, routinely over-claimed for their safety achievements (Adams, 1985). The problem is not confined to the road safety arena. Figure 7 by Leeth and Hale (2013) in their examination of the effect of the Occupational Safety and Health Act of 1970 suggests little has changed over the intervening decades. On their graph, displaying a downward trend similar to that in Figure 6, it is very difficult to discern the much-heralded effects of the Act.

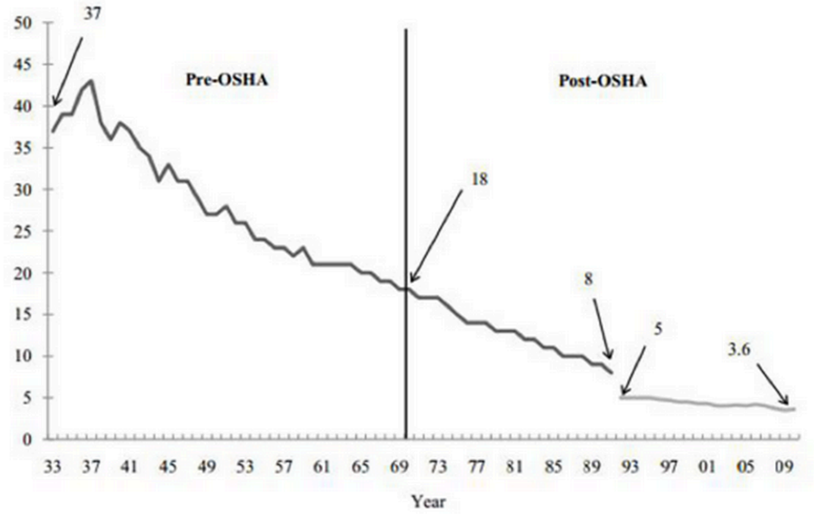


Figure 7. An evaluation OSHA's effectiveness
[Reproduced with the kind permission of the authors]

Almost 35 years ago Laurence Ross (1976), in the Scandinavian Myth, used a similar line-on-a-graph method to challenge the widely proclaimed view that Sweden, with its low permitted alcohol levels, strict enforcement, and draconian punishments was a model for the rest of the world to follow. His interrupted time-series analyses revealed no effect of the Scandinavian drink-drive laws on the relevant accident statistics.

His analysis suggested that tough drink-drive legislation is only likely to work where it accords with prevailing public opinion. He noted the existence of a politically powerful temperance tradition in Scandinavia. Many people considered drinking and driving a serious offence (if not a sin) before it was officially designated as such by legislators. The absence of a detectable effect of Scandinavian drink-drive laws on accident and fatality statistics at the time the laws came into effect suggested, according to Ross, that the laws were symptomatic of a widespread concern about the problem, and that most people likely to obey such laws were already obeying them before they were passed. The laws, in effect, simply ratified established public opinion.

I offer an examination of safety claims as a challenge to mathematical enthusiasts everywhere. It is a game the whole class can play. And it can be more than a game. It can be an introduction to the fascinating world of risk. There are vast numbers of risk management proposals and claims begging to be tested by mathematical enthusiasts. The risk-management “starter list” presented at the beginning of this essay merely scratches the surface.

In conclusion I return to risk on the road and invite others to share my fascination with the problems it presents. This is where I began over 40 years ago when challenging the Government’s road safety arguments at public inquiries and I still find some of the questions it raises challenging.

A final set of numbers – and a hypothesis

How might we account for the dramatic fall in numbers of those killed on the road as traffic increased since the Second World War in economically developed countries such as Britain? I offer Figure 8 as a basis for a hypothesis.

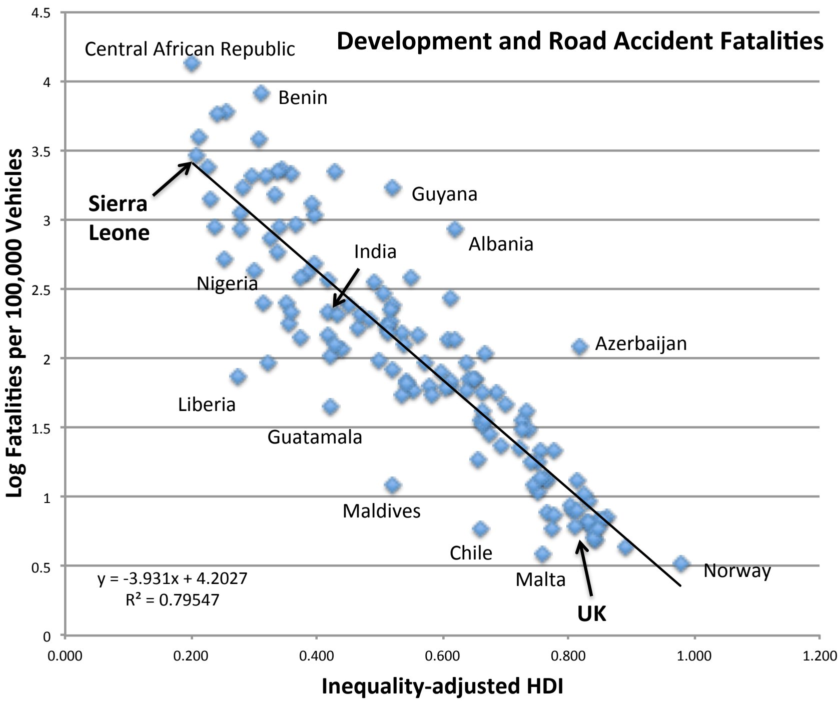


Figure 8. Development and road accident fatalities

The Central African Republic at the top of the top of the graph has a fatality rate per vehicle more than 3000 times higher than Norway at the bottom. And yet it has, along with most of the other countries at the top end, a full set of road safety laws: national speed limits, drink-drive limits, helmet laws, seat belt laws, child restraint laws and laws forbidding the use of mobile phones while driving. And they are all achieving their extraordinary kill rates per vehicle with modern imported vehicles with a hundred years of safety technology built into them. Norway's superior roads also appear unlikely to contribute to the difference; it is often remarked that potholes are nature's speed humps.

It appears that the process of "development" is accompanied by increased risk aversion and a growing sense of collective responsibility. A strong correlation exists between a country's score on the Equality Adjusted Human Development Index and its road death rate per vehicle. Created by Mahbub-ul-Haq and Nobel Laureate Amartya Sen, the Inequality-Adjusted Human Development index is a composite of average longevity, education and income, adjusted for inequality (http://en.wikipedia.org/wiki/Human_Development_Index). The largest outliers have been identified as a spur to further research.

For people living through the period represented by Figures 5 and 6 it would be difficult to perceive their roads getting 5.3% safer year on year. But that 62-year period has witnessed extraordinary societal change, and not just in the workplace as noted in Figure 7. As a child I can remember my respectable parents urging "one for the road" on departing guests. Now drunken driving has become a stigmatizing offence.

Over this period the freedom of children has been severely constrained. I grew up as a free-range child at liberty to roam the neighbourhood until the streetlights came on and expected to get to school on my own. A study of English schools in 1971 revealed that 80% of 7 and 8 year old children got to school on their own, unaccompanied by an adult. A follow-up study of the same schools in 1990 revealed that that number had fallen to 9% - and the main reason parents gave for denying their children the freedom that they had enjoyed as children was fear of traffic (Hillman, Adams & Whitelegg, 1990). And now it has become a legal child-protection issue. In England two controversies recently appeared in the press in which parents were threatened with child protection orders for allowing their children what used to be the widely accepted freedom to get to school unaccompanied (Moore & Maxted, 2010; BBC News, 2010).

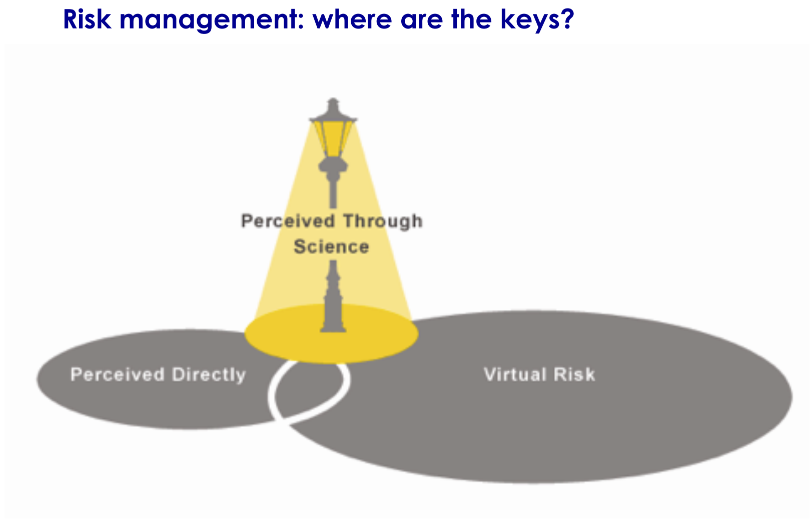


Figure 9. Risk management: where are the keys?

Figure 9 is a redrawn version of the Venn diagram of Figure 1, revised in homage to the mythical drunk who lost his keys in the dark and searched for them under the lamppost because that was where there was light to see. Risk managers searching for the keys to their problems amidst the brightly lit numbers in the mathematical circle are likely to be equally disappointed.

My wife and I usually buy a Euro Lottery ticket every week. The chance of winning is one in 116,531,800. My statistically minded friends jeer at me. My question to them is where else can we buy a week's worth of so much fantasy for £2? We spend an enjoyable week being wonderfully charitable to worthy causes and generous to friends and family – and enjoying a bit of self-indulgence. Our fantasies are highly improbable – but not impossible[2].

Even when risks can be calculated with great numerical precision the numbers can only be used to inform judgment, not substitute for it. And it matters greatly who is making the judgment.

References

Adams, J. (1985). *Risk and Regulation: the record of road safety regulation*. Retrieved from: http://john-adams.co.uk/wp-content/uploads/2007/10/risk%20and%20freedom.pdf

BBC News (2010, September 13). Lincolnshire family warned over girl's bu stop walk. Retrieved from: http://www.bbc.co.uk/news/uk-england-lincolnshire-11288967

Bloomberg News (2011). Risk officer rises to $10 million job after market meltdown. Retrieved from: http://www.bloomberg.com/news/2011-07-11/risk-officer-rises-to-10-million-job-after-derivatives-meltdown.html

Donaldson, C., Baker, R., Mason, H., Jones-Lee, M. W., Lancsar, E., Wildman, J., Bateman, I., Loomes, G., Robinson, A., Sugden, R. C., Prades, J. L. P., Ryan, M., Shackley, P. and Smith, R. (2011). The social value of a QALY : raising the bar or barring the raise? *BMC Health Services Research, 11*(8), 1-8.

[2] This example can serve as an illustration of the, often crucial, distinction between relative and absolute risk. In one recent draw the jackpot was £128 million. My wife bought two tickets. I stuck with one. Her "risk" of winning was twice that of mine [relative risk]. Her risk of winning was 1 in 58,265,900 times greater than mine [absolute risk].

Grossman, L. (2010). Metric math mistake muffed mars meteorology mission. Retrieved from: http://www.wired.com/2010/11/1110mars-climate-observer-report/

Hillman, M., Adams, J., & Whitelegg, J. (1990). *One false move...* Policy Studies Institute: London, England,

Leeth, J. & Hale, N. (2013). *Evaluating OSHA's effectiveness and suggestions for reform*. Retrieved from: http://mercatus.org/publication/evaluating-oshas-effectiveness-and-suggestions-reform

Moore, A. & Maxted, A. (2010). Should the Schonrock children be allowed to cycle to school alone. Retrieved from: http://www.telegraph.co.uk/women/mother-tongue/7872970/Should-the-Schonrock-children-be-allowed-to-cycle-to-school-alone.html

Ross H. L. (1976). The Scandinavian myth: the effectiveness of drinking and driving legislation in Sweden and Norway. *Evaluation Studies – Review Annual vol.*1, Sage.

Worth the Risk?
Modeling Irrational Gambling Behavior

Matt Lane
Math Goes Pop, Austin, Texas, USA

Abstract: In math class, expected value is often used when deciding whether or not a game is worth playing. A common refrain is that games with negative expected value should be avoided. However, nearly all games of chance have a negative expected value, and a simple expected value analysis fails to explain why these games are so popular. In this article, we consider three psychological factors leading to irrational gambling behavior – the illusion of control, hypersensitivity to reward, and beginner's luck – and explore how these factors affect an otherwise purely rational model of gambling behavior.

Keywords: expected value, gambling, roulette, luck, utility.

Introduction

Imagine I invite you to play a game. The rules are simple: each round, you have a 53% probability of losing a dollar (to me), and a 47% probability of winning a dollar (from me). We can play this game for as long as you like. Would you play?

If you're a math teacher, your answer is probably no. After all, this game, besides sounding as dull as can be, has a *negative expected value*: over time, you should expect to lose money and I should expect to gain it. Indeed, given these rules, your expected value for any one round is equal to:

$$0.53 \times (\text{-\$1}) + 0.47 \times (\text{+\$1}) = \text{-\$0.06}.$$

That is, on average, you can expect to lose 6 cents every time we play. Surely, there's no upside for you, and you'd be better off walking away.

And yet, these probabilities match some of the most common bets in a spin of the roulette wheel. The wheel has 18 red wedges and 18 black wedges, plus 2 green wedges, all of equal size. This means a bet on red (or black) has an $18/38 \approx 47\%$ probability of winning, and a $20/38 \approx 53\%$ probability of losing. But in spite of these numbers, people play roulette in droves. How silly of them; with but a little knowledge of probability, we decry in math class, people would know the risks and could save themselves considerable heartache.

This is the story students often hear when we talk about games of chance. If a game has negative expected value, you shouldn't play it; if it has positive expected value, you should. As with so many mathematical models, though, the reality is typically much more complicated. Offering games with positive expected value isn't exactly a viable long-term strategy for casinos, so most games with money on the line have a negative expected value. More importantly, people play games with negative expected value all the time, in spite of the risk involved. The interesting takeaway, especially from the standpoint of a classroom discussion, isn't: *people shouldn't do this.* Rather, it's: *why do people do this*? The answer to the latter question requires us to think more critically about our mathematical model, and try to adapt it more faithfully to our reality.

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 31-37

There are a number of reasons why people gamble, even though rationally they shouldn't. For example, the gambler's fallacy – the (incorrect) idea that if you've lost a number of times in a row, you're "due" for a win – is a common refrain. But there are subtler reasons that get touched upon less often, even though they can influence our decision-making in significant ways. In what follows, we'll touch upon three ideas: *illusion of control*, *hypersensitivity to reward*, and *beginner's luck*, and see how each of them affect our willingness to take on risk.

The Model

Since we'll return to it often, let's take a moment to abstract our model a bit. Instead of a game in which you win \$1 with a 47% probability, and lose \$1 with a 53% probability, we'll talk about a game in which you win an amount a_W with probability p, and lose and amount a_L with probability $1 - p$. For this to make sense, let's assume that a_W and a_L are both nonnegative. As before, your expected value in this game is equal to:

$$a_W\, p - a_L\,(1 - p) = (a_W + a_L)p - a_L.$$

If you believe that you shouldn't play a game with negative expected value, then you should only play this game if

$$(a_W + a_L)p - a_L \geq 0,$$

that is, if

$$p \geq \frac{a_L}{a_W + a_L}.$$

In particular, if the stakes are equal to the potential payout (that is, if $a_W = a_L$) you should only play the game if you have at least a 50% chance of winning.

People rarely follow this advice, however. Let's explore some reasons why.

Illusion of Control

The game I described above lacked any sort of context about *how* the winning and losing probabilities are determined. Is it by spinning a roulette wheel? Rolling dice? Playing with cards? Drawing marbles from a bag?

For the mathematically minded, these details may seem irrelevant. After all, once we know the probabilities, everything else is window dressing, at least in terms of our decision about whether or not to play.

But in reality, context is extremely important. This is due, in part, to the *illusion of control*, i.e., the belief that people have much more control over a random outcome than they actually do. When people believe they control over the outcome of a game, they are likely to take greater risks, even if objectively they have no control at all.

There are several ways to elicit the illusion of control in a game of chance. One way is to give the player something to do, or some decision to make, even if they have no control over the outcome of what they're doing. In roulette, players can choose what to bet on: whether the ball will stop on red or black, an even number or and odd number, and so on. In dice games, players roll the dice. Even though the player has no control over how the dice will land, simply giving the player the opportunity to roll can elicit the illusion of control. This manifests itself in subconscious ways: for example, an experiment by

Henslin (1967) found that when players needed to roll low numbers in a dice game, they rolled the dice more softly, and when they needed high numbers, they rolled more forcefully!

Take away the activity granting players an illusion of control, and their confidence will waver as a result. Strickland, Lewicki and Katz (1966) illustrated this point with an experiment: they separated students into two groups, both of which played a dice game. In one group, however, they players bet before they rolled the dice; in the other group, the players bet *after* they rolled the dice (but before they knew the results of the roll). The experiment revealed that people who bet after rolling the dice tended to be more conservative in their wagers than people who bet beforehand.

Players also experience more illusion of control when the game itself is familiar. Burger (1986) found that people with a "high desire for control" will bet more on a card game when playing with a familiar deck of cards (using the standard club, spade, heart and diamond symbols) than when playing with cards that used unfamiliar symbols. Of course, when it comes to a random process, familiarity makes no difference in one's ability to predict the outcome.

The illusion of control is great for casinos, but not so great for individual players. And the problems that arise from this illusion are compounded by the fact that mathematically, most casino games operate on a relatively small negative expected value. The illusion of control isn't likely to make you think you should pay $90 to play a game where you have only a 1% chance of winning $100. But typically the house edge is quite small, and just a tiny nudge in your perceived probability of winning may make you willing to take a risk that you otherwise wouldn't.

Put another way, for any given casino game, it likely isn't true that $p \geq \frac{a_L}{a_W + a_L}$ holds. However, if you believe your probability of winning is actually some value $p^* > p$, it may be true that $p^* \geq \frac{a_L}{a_W + a_L}$, in which case you may play the game in spite of the risk involved. When betting a dollar on black in roulette, for instance, $p \approx 47\%$, and $\frac{a_L}{a_W + a_L} = 50\%$. So if the illusion of control can bump up your confidence in winning by just three or four percentage points, you may opt to put your money on a spin of the wheel. Unfortunately, the *illusion* of control is just that, and while confidence is an asset in many aspects of life, a little humility when gambling is probably wise.

Hypersensitivity to Reward

Most analyses of games of chance are framed in terms of dollars: how much will you make if you win, and how much will you lose if you don't? But for many people – compulsive gamblers especially – this isn't necessarily the right perspective. Sometimes, people don't gamble because they think they can win. Rather, they gamble because it feels *good* to win.

Instead of talking about dollars, then, maybe it makes sense to talk about gambling in terms of something more abstract, like *utility*, which serves as a proxy for measuring happiness. Unlike the illusion of control, which affects the *probabilities* of winning, swapping dollars for utilities affects the *payouts* if we win or lose.

Notationally, not much changes. If you gain a utility u_W by winning and lose a utility u_L by losing, then we should only play a game of chance if $p \geq \frac{u_L}{u_W + u_L}$. But this leaves us with a couple of questions:

1. How do utilities compare to dollar amounts?
2. How do utilities vary among people?

The first question is difficult to quantify. At first, it may seem reasonable that utility should be roughly proportional to dollar amounts. If you win twice as much, you should be twice as happy; winning \$10 should bring you as much happiness as losing \$10 costs you. And so on.

However, this doesn't make for much of an interesting story, since if people's utilities were proportional to amounts won and lost, the ratios $\frac{u_L}{u_W + u_L}$ and $\frac{a_L}{a_W + a_L}$ would be the same, and we'd be no better off trying to explain people's behavior by using utility instead of money.

Nevertheless, it may be true that for *some* people, utility is roughly proportional to actual amounts wagered. Unlike dollar amounts, though, which are objective and uniform, there's likely much more variation among individuals when it comes to utility fluctuations while gambling.

Which brings us to our second question. Utility variation among individuals is of particular interest when it comes to two groups: compulsive gamblers and everyone else. People who have a gambling problem will dig themselves into deeper and deeper financial holes, taking on financial risks well past the point at which any reasonable person would walk away (or so we like to tell ourselves). Illusion of control may play a part in this, but utility plays a role as well.

There are two possible explanations for how a change in utility could lead to riskier gambling behavior: either a person doesn't feel the sting of losing (u_L is relatively low), or the person feels a heightened euphoria after winning (u_W is relatively high). Of these options, the work of Hewig et. al. (2010) has shown that the second is more plausible. By comparing the electrophysiologic responses of problem gamblers to a control group, their experiment showed that problem gamblers are hypersensitive to reward, possibly due to an above-average increase in dopamine when they win.

In other words, the factor u_W may be higher for a problem gambler than for a non-problem gambler. But as u_W increases, the ratio $\frac{u_L}{u_W + u_L}$ decreases, meaning that a problem gambler may be more willing to risk playing a game with relatively poor odds.

For instance, suppose there are two people thinking of playing a game. One person's utility is proportional to the dollar amounts she can win or lose, while the other is hypersensitive to reward. Their utilities may look something like this:

	Person A	**Person B (Hypersensitive)**
Utility if Win	u_W	$1.5u_W$
Utility if Lose	u_L	u_L

Now suppose the potential payoff in the game equals the potential loss ($a_W = a_L$). Since Person A's utilities are proportional to the dollar amounts, this means that

$$\frac{u_L}{u_W + u_L} = \frac{a_L}{a_W + a_L} = 0.5,$$

so Person A will only play the game if her probability of winning is at least 50%. But for person B, the threshold is lower:

$$\frac{u_L}{1.5u_W + u_L} = \frac{a_L}{1.5a_W + a_L} = 0.4.$$

Returning again to the example we began with, Person A won't play this game if the probability of winning is 47%, but Person B will!

Framed in terms of money, both people have the same expected value for this game: -0.06a_W. For Person A, this is proportional to their expected utility from playing the game: -0.06u_W. However, Person B's expected utility is equal to 0.175u_W. Person B expects to gain utility from this game, but lose money. (Of course, utility and money aren't independent, and even someone who is hypersensitive to reward will likely start to feel bad after losing a lot of money in one sitting.)

Combined with the illusion of control, hypersensitivity to reward can be even more dangerous. The illusion of control inflates our perceived probability of success (from p to p*), while hypersensitivity to reward inflates the value of a win (from a_W to u_W). Both of these factors increase our willingness to take on risk. Taken together, instead of checking the inequality

$$p \geq \frac{a_L}{a_W + a_L},$$

for some people, the relevant inequality is

$$p^* \geq \frac{u_L}{u_W + u_L}.$$

When $p^* > p$ and $\frac{a_L}{a_W + a_L} > \frac{u_L}{u_W + u_L}$, it's quite possible for the second inequality to hold even if the first one doesn't.

Beginner's Luck

We've seen that illusions of control and hypersensitivity to reward can skew our perceptions of games of chance. Research also suggests that some people are more susceptible to these phenomena than others. But before you breath a sigh of relief, convinced of your resistance to these risky behaviors, you should know that, given the right circumstances, none of us seems completely immune.

For example, in one of the more famous papers on the illusion of control, Langer and Roth (1975) found that *beginner's luck* has a measureable affect on people's perceptions of success. One important aspect of the experiment was that the subjects weren't control freaks or problem gamblers; they were simply undergraduates at Yale.

In the experiment, 90 male subjects were asked to predict the outcome of a coin flip thirty times in a row. While the experiment was perceived to be random, subjects were actually put into one of three groups. Members of one group (the *descending* group) were told they guessed correctly very frequently early on, but less frequently as the game progressed. Members of another group (the *ascending* group) were told the opposite: many incorrect guesses early on, but more correct guesses later. A third group was given a randomly generated string of wins and losses to serve as a control.

Descending Group	WWWWLWWWLLLWWWLLWLLWWLLLWLLLLW
Ascending Group	WLLLLWLLLWWLLWLLWWWLLLWWWLWWWW
Random Group	WLWLLWLWLWWLLWLLWLWWWLWLWLLWLW

The members of each group were told they won 15 times and lost 15 times; the only difference between groups was the way in which the wins and losses were arranged. Afterwards, each subject was asked a few questions about his performance. Here are some of the questions, along with the average response for each group:

	Descending	Ascending	Random
From 0 to 10, how good do you think you are at predicting outcomes like these?	5.70	4.93	5.10
How many times did you correctly guess the coin toss?	16.7	13.3	14.4
After 100 more rounds, how many more correct guesses do you think you'd make?	54.2	49.1	51.1

People who started off with more wins believed they were better at predicting coin flips than people in the other two groups (the illusion of control rears its head once again). Also, even though people in each group "won" exactly 15 times, people in the descending group thought they had won more than 15, and people in the ascending group thought they had won fewer. Finally, when asked about their probability of future success, people in the descending group had a higher estimate (54.1%) than people in the ascending group (49.1%).

In other words, early success when playing a game can influence feelings of control, even among people who are ostensibly highly rational. A person who wins k rounds out of a total of n rounds in a game of chance should estimate their probability of winning a future round by dividing the number of wins (k) by the number of games played (n):

$$\frac{\#\{\text{wins}\}}{\#\{\text{games played}\}} = \frac{k}{n}.$$

In reality, however it seems that people tend to give more credence to wins that occur earlier on.

We can attempt to model this mathematically as well. Suppose a player weighs the n^{th} round of a game according to some weighting function $w(n)$. If the player wins k rounds i_1, i_2, … i_k out of a possible n, then he will estimate his probability of future success by the ratio

$$\frac{w(i_1) + w(i_2) + \ldots + w(i_k)}{w(1) + w(2) + \ldots + w(n)}.$$

A perfectly rational person will weigh the importance of each round equally, in which case we can take w to be the constant function 1, and the fraction above reduces to k/n – that is, the number of wins divided by the number of plays. For someone who values early wins more highly, however, w is likely to not be a constant function. Instead, it seems reasonable to insist on the following two properties:

1. w is non-increasing,
2. $w(n) \to 1$ as $n \to \infty$.

The first assumption ensures that we really do weigh initial rounds more than (or at least, not less than) later rounds. If the second assumption holds, we become more rational as time progresses; that is, eventually, all rounds are weighed roughly the same.

For instance, suppose we give the n^{th} round a $(100/n)\%$ significance bonus, i.e. we take $w(n) = 1 + 1/n$. This satisfies both of the assumptions above, and compared to the rule $w(n) = 1$, we see that it does indeed favor earlier rounds to later rounds. We can also apply this rule to the ordering of wins and losses in each group listed above in order to estimate the perceived probabilities of future success for each group:

	Descending	Ascending	Random
Weighted probability ($w(n) = 1 + 1/n$)	52.9%	49.6%	50.7%
Self-reported probability (from above)	54.2%	49.1%	51.1%

This relatively simple weight function does a reasonable job predicting the self-reported outcomes, and no doubt a more sophisticated model could do even better. Regardless of the model, the point remains: even for people with no history of gambling or control issues, just a little bit of beginner's luck can be enough to get people to engage in risky gambling behavior. So if you do decide to gamble, maybe the best thing for your wallet is a string of early losses.

Conclusion

Talking about gambling in the classroom can be tricky, but the moral of the story is usually one that everyone can get behind. The math says you shouldn't gamble, because you can expect (in the mathematical sense) to lose. However, this surface-level analysis doesn't do much to explain why people *do* gamble.

One approach to consider is that while the expected value on most gambling events is negative, it's frequently only slightly negative. In this case, it doesn't take much psychological persuasion to convince people that their odds seem better than they objectively are. The illusion of control, hypersensitivity to reward, and beginner's luck are three ideas that can help explain why people take on more risk than they should, and all three can be incorporated into mathematical models that use expected value. However, these ideas are rarely explored in the classroom. But in the end, if we want to discourage students from taking irrational risks with their money, it seems better to try to account for that irrationality rather than to ignore it.

References

Burger, J. M. (1986). Desire for control and the illusion of control: the effects of familiarity and sequence of outcomes. *Journal of Research in Personality, 20,* 66-76.

Henslin, J. M. (1967). Craps and magic. *American Journal of Sociology, 73,* 316-330.

Hewig, J. et al. (2010). Hypersensitivity to reward in problem gamblers. *Biological Psychiatry, 67,* 781-783.

Langer, E. J. and Roth, J. (1975). Heads I win, tails it's chance: the illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology, 32(6)*, 951-955.

Strickland, L. H., Lewicki, R. J., Katz, A. M. (1966). Temporal orientation and perceived control as determinants of risk-taking. *Journal of Experimental Social Psychology, 2,* 143-151.

Understanding Risk Through Board Games

Joshua T. Hertel
University of Wisconsin–La Crosse, USA

Abstract: In this article, I describe a potential avenue for investigating individual's understanding of and reactions to risk using the medium of board games. I first discuss some challenges that researchers face in studying risk situations. Connecting to the existing probabilistic reasoning literature, I then present a rationale for using board games to model these situations. Following this, I draw upon intuition and dual-process theory to outline an integrated theoretical perspective for such investigations. The article concludes with two vignettes demonstrating how this perspective might be used to analyze thinking about risk in a board game setting.

Keywords: risk, probabilistic reasoning, intuition, dual-process theory, board games.

Within our information-driven society, understanding and making decisions about risk has become important for personal and professional success. Each day individuals are bombarded by information from a host of organizations seeking to inform, influence, and persuade choices through the use of hypothetical situations involving risk, which I refer to as risk-laden (RL) situations. Here RL situations are defined as any situation that presents a perceived risk directly (e.g., men over the age of 25 have increased risk for gestational diabetes) or presents a risk indirectly by using probabilistic language (e.g., the odds of dying from heart disease are 1 in 7). Adults, in particular, are targeted by a host of groups (e.g., news organizations, advertisers, health professionals, financial planners, insurance companies) each of whom seeks to use RL situations to inform, influence, and motivate choices.

RL situations can be differentiated using two criteria, which I will call (and later define) density and complexity. Density refers to the extent to which probabilistic information is incorporated into a RL situation. Low-density RL situations present simple probabilities whereas high-density situations use sophisticated probabilistic language that may be embedded or indirectly expressed. Complexity refers to the context of an RL situation. Everyday contexts have a relatively low complexity whereas specialized contexts have high complexity. These criteria are useful in comparing RL situations directly. For example, consider the following two situations: (a) A meteorologist reports that there is a 60% chance of rain this weekend; and, (b) Doctor John shows Kim the following finding from a medical study, "When used in primary prevention settings, aspirin has been shown to reduce serious vascular events among individuals at average/low risk by 12% (0.51% versus 0.57%/year, P = 0.0001)" (Cuzick et al., 2014, p. 5). The first situation has a low degree of density and complexity because it presents a simple probability within an everyday context. The second has a higher degree of complexity and density because it uses sophisticated mathematical language set within a more specific context. Thus, RL situations with high degrees of density and complexity embed sophisticated probabilistic information within specialized contexts.

Although different in terms of density and complexity, both of the previously mentioned situations present similar issues in decoding. For example, what does a 60% chance of rain mean? Does each day during the weekend have a 60% chance of rain? If it rains the first day, will the next day still have the same chance of rain? Similarly, what does it mean for an individual to be at average/low risk? What information factors into this classification? What does a 12% reduction in serious vascular events

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 38-54

amount to in terms of overall risk for these events? Additionally, the interpretation of each statement might change based upon the person making it. What if the meteorologist in the first situation worked for the National Weather Service? What if they worked for an insurance company? What if Dr. John were Kim's family doctor? What if he worked as a consultant for a retailer? These examples illustrate that, regardless of density and complexity, RL situations present similar challenges to individuals in terms of unpacking and understanding.

Although mathematics education researchers have investigated a range of probabilistic concepts, RL situations present different challenges for several reasons. First, RL situations often require one to make sense of several different hypothetical outcomes and weigh competing possibilities against one another in order to make sense of the situation and determine the most beneficial decision. In some instances the best decision might yield a beneficial outcome, but in others the best decision might be the outcome with the least potential consequences. Second, in contrast to familiar contexts where the probability of an event does not change (e.g., rolling a six on a fair die), RL situations can be dynamic with the likelihood of outcomes changing as events unfold. For example, weather events can change the potential risk for flooding, shifts in economic conditions influence the risk related to financial instruments, and a recent health issue can affect the risk of a medical procedure. Third, the range of contexts for RL situations can present an obstacle since individuals may have limited background knowledge about a context and must instead rely on probabilistic knowledge, experience, and intuition to make a decision.

Taken together, these points suggest that if mathematics educators wish to pursue research on mathematical thinking about risk, the setting of these investigations should be flexible with the capability of addressing both the density and complexity of RL situations. This article describes one possible avenue using the dynamic and diverse medium of board games. In what follows, I first discuss mathematics education research literature on probabilistic thinking and draw connections to RL situations. I then define what is meant by the phrase board game and present a rationale for using the medium of board games to model RL situations. Following this, a theoretical perspective is presented that may offer assistance in analyzing RL situations. The article concludes with sample vignettes modeling how the perspective might be used within a board game setting to understand an individual's thinking about risk.

Research on Probabilistic Reasoning

Probabilistic reasoning has been a focus of researchers in both mathematics education and psychology (Chernoff & Sriraman, 2014) and presents a rich foundation from which to build investigations of individual's thinking about risk. Central in probabilistic reasoning is an individual's understanding of randomness. As noted by Batanero, Green, and Serrano (1998) randomness resides at the heart of probabilistic reasoning because it serves as a string that binds together a collection of different mathematical concepts. Although there are many interpretations of randomness, the present work uses a definition presented by Moore (1990),

> Phenomena having uncertain individual outcomes but a regular pattern of outcomes in many repetitions are called *random.* "Random" is not a synonym for "haphazard" but a description of a kind of order different from the deterministic one that is popularly associated with science and mathematics. Probability is the branch of mathematics that describes randomness. (p. 98)

Based upon this definition, although the outcome of a specific event itself may be uncertain and unpredictable, if the same event is repeated a large number of times, patterns emerge and yield frequencies that, in turn, make prediction possible (Metz, 1997). Randomness connects a cluster of related ideas including uncertainty, likelihood, and chance. Therefore, an individual's probabilistic reasoning is built upon a foundation of knowledge of randomness and this collection of related ideas.

Research has documented a variety different non-normative (i.e. not held by experts) ideas that individuals may hold about randomness (Ayton & Fischer, 2004; Gold, 1997; Shaughnessy, 1992). One well-known example is the gambler's fallacy. Roughly speaking, this is the belief that a series of one outcome will create a tendency for another, opposite outcome. People who hold the gambler's fallacy often believe that they can predict the next outcome of a random process based on prior observations. For example, a coin is flipped six times resulting in the sequence THTTTT. What will be the outcome of the next flip? A person holding the gambler's fallacy would likely predict that the next outcome would be heads. This is because they believe that the observed sequence of tails would need to be balanced by a sequence of heads. However, from a normative viewpoint both heads and tails are equally likely.

Furthermore, the idea that random processes have a "bookkeeping" ability (i.e. the process remembers and reacts to previous results) is not isolated to typical probability situations. This is illustrated by the following Dear Abby letter:

> Dear Abby: My husband and I just had our eighth child. Another girl, and I am really one disappointed woman. I suppose I should thank God that she was healthy, but, Abby, this one was supposed to have been a boy. Even the doctor told me the law of averages were in our favor 100 to 1. (Dawes, 1988, p. 84)

Although the probability that a particular couple will have eight girls is quite small (roughly 0.0039 if we assume that a boy and girl are equally likely), the fact that a couple has already had seven girls does not change the probability that the next child will be a girl (it is still .5). However, as is evident in the excerpt, both the woman and her doctor believed that the next child would have to be a boy by the "law of averages." This statement is an application of the gambler's fallacy since both the woman and her doctor have assumed that the random process will "even out."

The previous Dear Abby example illustrates an issue concerning probabilistic knowledge. Unlike many other types of mathematical knowledge, which are encountered almost entirely within classroom settings, probability situations are encountered frequently as individuals go about daily activities outside of school. This is because the stream of RL situations within our modern society has become a constant part of communication. At the same time, the role of probability within the pK–12 curriculum has remained relatively minor within the United States. This is reflected in the limited emphasis of probability within current content standards (Mooney, Langrall, & Hertel, 2014). Thus, probabilistic knowledge is learned via in-school and out-of-school experiences; however, as I will discuss, out-of-school experiences do not always offer information that supports decision-making in RL situations.

One issue with out-of-school experiences is that individuals can easily be led astray by culturally accepted ideas that persist in a variety of different formats (e.g., maxims, epigrams, anecdotes, fables, proverbial sayings). Many of these ideas are encoded with probabilistic information, which can influence an individual's reasoning. For example, consider the adage "lightning never strikes twice." If a person truly believes that lightning will never strike the same place twice, then they are more likely to stand beneath an object that has previously been struck by lightening during a thunderstorm. In reality this belief is unfounded since lightning can, and does, strike the same location multiple times. Uman (1986) noted that "the Empire State Building is struck by lightning an average of about 23 times per year. As many as 48 strikes have been recorded in one year, and during one thunderstorm, eight strikes occurred within 24 minutes" (p. 47).

Another issue with out-of-school experiences is that they may teach individuals to focus on less important aspects of a RL situation. For example, a focus solely on the probability of lighting striking at the same location obfuscates other important information about the context. Since risk is intimately tied to context, the context should be considered when reasoning through RL situations (e.g., one should consider the physical landscape when making a decision about where to seek shelter in a lightening storm). Additionally, out-of-school experiences may promote a context-free application of ideas that can lead individuals astray. For example, the previous adage can be applied in relevant contexts (e.g., lightening storms) as well as other situations that an individual deems appropriate (e.g., winning the lottery), thereby obstructing and derailing decision making about unrelated events.

As mathematics education researchers, our primary goal is to investigate the learning and teaching of mathematical ideas. Although the focus of the field has tended to be on the teaching and learning of mathematics of children and young adults within school settings, there are several reasons that these conditions are constraining for studying probabilistic reasoning and its influence on decisions in RL situations. First, as previously noted, probability has a relatively weak position within most current curricula. This means that the majority of students have little in-school experience with probabilistic situations. Second, individuals spend only a short span of years in school but are subjected to the constant presence of RL information throughout their lifetimes. Although out-of-school experiences may provide individuals with some additional probabilistic knowledge, these experiences may also hinder individuals by reinforcing ideas that lack the details and specifics needed to assist them in making rational decisions about RL situations (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993). Thus, these out-of-school experiences have the potential to hinder or impede reasoning.

Third, the majority of existing research has used contrived situations crafted in contexts outside of participants' real life experience (e.g., Piaget's tip box, the hospital problem). Although this work provides insight into probabilistic reasoning, it offers little information how this reasoning is applied in more familiar contexts. Missing are studies that investigate participants' probabilistic reasoning within a context that is more closely related to an individual's normal routine. These are the RL situations that individuals are continually faced with and must make decisions about.

Fourth, viewed as a whole, the primary goal of past research has been to identify and study particular concepts or misconceptions, but little is known about how these ideas impact more complicated decisions. Studies on development of probability concepts have primarily sought to establish when particular concepts develop and if it is possible to remediate misconceptions. Judgment and decision-making studies have investigated the reasons behind the choices that people make and sought to understand commonly held misconceptions (e.g. the gambler's fallacy, the hot hand belief). Likewise, research investigating probabilistic intuitions directly has yielded mixed results with some intuitively based probabilistic misconceptions found to weaken with age, others found to grow stronger, and still others found to stabilize (Fischbein & Schnarch, 1997). Overall, missing from the literature is research that examines how these factors combine to influence decision making *in action*.

Finally, RL situations present challenges for interpretation that are different from classic probability problems. This is because RL situations are often presented as facts to be consumed rather than predictions based upon the observed frequencies of a random phenomenon. As a result, the assumptions and limitations of a prediction, which tend to be evident in more traditional probability problems, can be easily lost in communication about RL situations. The context-specific nature of risk means this poor communication can lead to misinformation and misunderstanding. Taking a previous example, it is impossible to predict with absolute certainty if it will rain today, but given that it has rained 60 out of the last 100 days with similar environmental conditions we can predict a 60% chance of rain. Thus, the risk of rain that is reported on a particular day is based upon a specific geographic location and, as experienced individuals can attest, a change in geography can invalidate this prediction. However, the RL information, which is shared and consumed, omits these specifics. This is quite different from more traditional educational settings where the context and assumptions are known or specifically focused on.

The Medium of the Board Game

For the purposes of this study, the phrase *board game* refers to a type of game that has the following characteristics: (a) a board, play matt, or clearly defined play area on a table or similar surface on which game pieces are placed and interacted with; (b) pieces, cards, or markers that are used for a variety of different purposes; (c) an external process or device that incorporates uncertainty into game play (e.g., spinner, dice, random card draw); and, (d) the absence of gambling with real-world currency (although wagering of in-game currency may be a component of gameplay). Popular games that fall under this definition include *Monopoly, Risk, Trivial Pursuit, Clue, Life,* and *Chutes and Ladders*. This definition excludes deterministic games such as chess or checkers, which do not have external processes that incorporate uncertainty into gameplay. Likewise, gambling-focused card games such as poker or blackjack are excluded from this definition.

Because board games contain external processes that incorporate uncertainty, players engage in probabilistic reasoning as part of normal gameplay. Moreover, the incorporation of uncertainty results in situations where players must make decisions based upon perceived risks and rewards. To illustrate a simple example, choices in the popular board game *Monopoly* focus on acquiring or selling properties. As the game progresses, players must make decisions about purchasing, selling, or improving properties. Since the number of spaces a player moves each turn is dependent on a dice roll, the number of turns it will take to travel the entire board, which is a primary means of collecting income, is uncertain. Consequently, the number of opportunities that a particular player will have to land on and purchase a given property is unknown. As the game progresses, opponents may benefit from acquiring properties, drawing random cards, or rent-free trips around the game board. On the other hand, opponents may be disadvantaged by paying the luxury tax, being sent to jail, or having to mortgage properties. Thus, the system of the board game, which includes the monetary assets of players as well as their property ownership, is dynamic. Players must continually reassess their options and adjust their actions as they play the game. Consequently, gameplay requires balancing the risks of running out of money with the rewards of acquiring property.

Monopoly has a relatively simplistic design when compared to many other board games because the driving force behind in-game events is the result of a dice roll. The decisions made by players about purchasing, selling, improving, or mortgaging properties have little influence on the consequences resulting from the dice roll. Instead, these decisions are mostly in reaction to random events. In contrast, many other board games provide players with a variety of choices during gameplay that can be made in anticipation of in-game events. This makes it possible for players to reduce the effect of random events or change the consequences of a particular event. By providing players with more choices, these more complex board games also provide opportunities to develop short-term and long-term strategies for managing risk.

The board game *Settlers of Catan* serves as one example highlighting these ideas. The game, which was developed in the 1990s, has grown greatly in popularity over the last decade. The objective of the game is to be the first person to accumulate 10 points. These points are gained by building structures, purchasing cards, or being awarded one of two special cards. The game board is made up of 19 hexagon tiles, which are arranged together in a specific pattern. One tile denotes the desert and the other 18 are one of five land types (hills, pasture, mountains, fields, forest). A special token (the robber) is placed on the desert and each of the other tiles is assigned one of the following numbers: 2, 3, 3, 4, 4, 5, 5, 6, 6, 8, 8, 9, 9, 10, 10, 11, 11, 12 (note that the number 7 is excluded from this list). Each of these numbers corresponds to a sum that can be obtained by rolling two six-sided dice. Figure 1 shows a portion of the board arranged in the recommended starting setup for beginners.

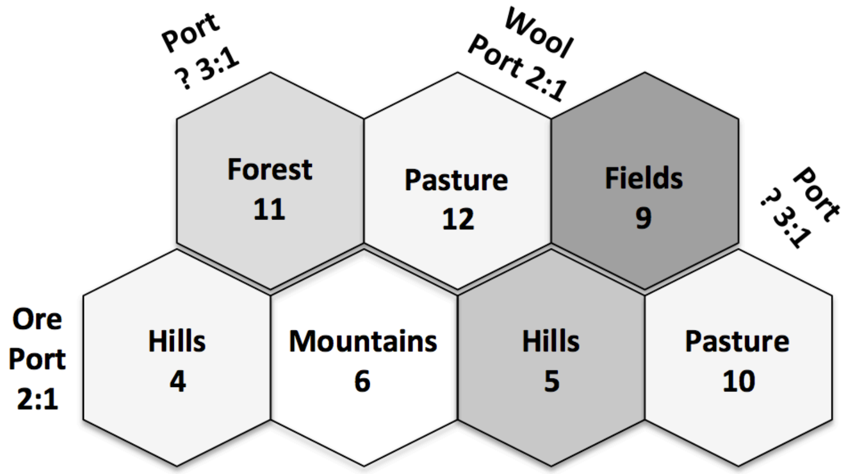


Figure 1. A portion of the *Settlers of Catan* game board

Like *Monopoly,* in-game events in *Settlers of Catan* are driven by the roll of two six-sided dice. Each turn the dice are rolled and any tile with the corresponding sum produces a resource (e.g., brick, wool, ore, grain, lumber). The most likely sum, 7, causes the robber to move to a different tile, which prevents the tile from producing resources and triggers other in-game events. Players collect resources by building structures (towns or cities) at the vertex of a tile. Thus, a player might build a town at a vertex shared by multiple tiles or choose to build at a vertex on the edge of the game board that is on a single tile. As shown in Figure 1, there are also ports along the edge of the board that, if built upon, allow a player to trade resources. For example, the port labeled ?3:1 allows players to trade three of one resource for one of another. Players also have the option to trade resources with each other or with the bank.

In contrast to *Monopoly, Settlers of Catan* offers a number of in-game choices that can change the potential risk (or reward) of a dice roll. Since gameplay is driven by collecting and using resources to acquire points, the most influential of these choices is arguably the initial placement of towns during the setup of the game. Initial placement limits future building because players must connect new structures to existing structures they control. During the game setup, players are given the initial choice to build two towns anywhere on the game board following specific in-game rules (e.g., each town is at least two edges from another town). There are a variety of strategies that players may adopt. For example, a player may choose to build at a specific vertex to collect a single type of resource. In Figure 1 a player using this strategy might build at a vertex of each hill tile to collect brick. This is likely to provide the player with a large quantity of bricks over the course of the game, which they can trade with the bank or with other players. On the other hand, a player may choose to build on tiles that have specific numbers. In Figure 1, the six on the mountain tile is the most likely sum to be rolled, followed by the five and nine. A player using this strategy may choose to build on the vertex shared by the Mountains, Pasture, and Hills thinking that two of these tiles have a greater likelihood of producing resources. These strategies, which are just two of many potential strategies, illustrate how the decisions made in playing the game are based upon both probabilistic reasoning and expectations about risk.

As the game progresses, players continue to build towns. However, these decisions must take into account the locations of opponents and the resources that a player has on hand. As areas of the game board fill with towns, the risk of not acquiring specific resources increases. At the same time, players are subject to the uncertainty of the dice roll, which can reward or penalize players in unforeseen ways. Likewise, players may choose to work together with opponents for mutual benefit or to prevent another person from acquiring points. Thus, the decisions that players make during a game offer insight into their in-action reasoning.

Although the proceeding discussion has focused on a seemingly complicated game, *Settlers of Catan*, the game has less complexity than other board games. In fact, within many board gaming circles *Settlers of Catan* is seen as a gateway game that is good for introducing new players to the medium of the board game. However, *Settlers of Catan* is often criticized for the relatively simple choices that players are allowed to make and the heavy reliance on the roll of dice.

Monopoly and *Settlers of Catan* can be seen as representing two initial points along a continuum of board games. This continuum differentiates games based upon two criteria, which I will call fortuity and intricacy. Fortuity refers to the extent to which probabilistic elements have been incorporated into a game. Board games with a low degree of fortuity incorporate a single random process or device (e.g., a dice roll each turn) whereas games with greater fortuity integrate several random processes. Intricacy is a measure of the decisions available to a player at any given point in the game. Board games that provide the player with a few simple choices have low intricacy while those that provide a large number of complicated choices have a greater degree of intricacy. Although fortuity and intricacy are independent of each other (e.g., it is possible to have a high fortuity and low intricacy), they are often linked. In particular, as board games increase in fortuity and more random processes are incorporated into gameplay, there is a tendency to increase intricacy thereby providing players with more choices to consider when making decisions. Both *Monopoly* and *Settlers of Catan* have relatively low fortuity because they incorporate simplistic random processes (e.g., dice rolls, random card draws) and present these processes clearly to players rather than embedding them into game mechanics. On the other hand, *Settlers of Catan* has a greater degree of intricacy than *Monopoly* because it offers players a variety of decisions during the game.

In addition to providing a method to categorize board games, the preceding criteria also furnish a means to connect board games to RL situations. The extent to which a board game can effectively model a given RL situation is related to the game's fortuity and intricacy. Games with a low degree of intricacy and fortuity are more suited to modeling RL situations with low density and complexity. These games have relatively simple rule sets, can be played quickly (often under an hour), and require minimal background knowledge. Consequently, they are well suited to modeling RL situations that present simple probabilistic information (low density) within a general context (low complexity).

On the other hand, board games with a high degree of fortuity and intricacy are able to model RL situations with a high degree of density and complexity. These games feature complicated rule sets, take longer amounts of time to play (four to six hours is typical), and require more extensive background knowledge. This allows the games to model RL situations with sophisticated probabilistic language (high density) that has been embedded into a specialized context (high complexity). Thus, these criteria provide a means to carefully select board games so that they closely model RL situations and provide a medium in which to investigate individual's thinking about risk.

Outlining a Theoretical Perspective

This article now offers one potential theoretical perspective for investigating RL situations using the medium of board games. This perspective is formed from the integration of two different theories on learning: intuition and dual-process theory. In what follows, I draw on Efraim Fischbein's work on the role of intuition in mathematical learning (Fischbein, 1987; Fischbein, & Gazit, 1984; Fischbein, Nello, & Marino, 1991; Fischbein, & Schnarch, 1997) and dual-process theory as describe by Kahneman (2002) and Leron and Hazzan (2006, 2009). Following a discussion of each theory, I then consider how they can be combined to form an integrated theoretical perspective for future studies.

Intuition

Fischbein's (1987) theory of intuition posits that intuitions have a cognitive-behavioral function. Fischbein defined an intuition to be "an *idea* which possesses the two fundamental properties of a concrete, objectively-given reality; *immediacy*—that is to say intrinsic *evidence*—and *certitude* (not formal conventional certitude, but practically meaningful, immanent certitude)" (Fischbein, 1987, p. 21). In other words, an intuition, or intuitive belief, is one that is immediately available to a person in a given situation and is held by the person (at least while they are expressing it) as being true. From this perspective, intuitions serve a similar behavioral function as perceptions in helping to guide our mental or practical activity. Fischbein argued that when faced with uncertainty or lack of information individuals naturally extrapolate beyond what they are given in order to make a decision. This is not, he contended, a unique event in one's existence. Rather, intuitions are part of our every-day experience. Fischbein noted,

> Promptly adjusted, well adapted reactions of a person to given circumstances are possible only if the perception of the respective reality appears to him, automatically, as coinciding with reality itself. Doubts, hesitations are useful only when referring to aims which are not directly involved in the current flow of behavior. When crossing the street, you have to believe absolutely in what you see—the approaching cars, the various distances etc.—otherwise your reactions will be discontinuous and maladjusted. Analogically, during a reasoning process, you have to believe—at least temporarily (but absolutely)—in your representations, interpretations or momentary solutions, otherwise your flow of thoughts would be paralyzed. It is this type of belief that we call an intuition. Cognitive beliefs, elaborated and confirmed repeatedly by practice, may acquire an axiomatic character." (p. 28)

Thus, Fischbein argued that intuitions reveal themselves when individuals are faced with making decisions. This is relevant to the setting of a board game, which routinely places players in situations where they must make choices in order to allow play to progress. These decisions rest in part upon their intuitions. Moreover, as Fischbein noted, intuitive beliefs can become axiomatic if they are reinforced by repeated practice. Repeated practice is common in board games where players must make the same or related decisions many times throughout the course of the game.

Fischbein saw intuitions as an essential aspect of human cognition and stressed the difference between a perception and an intuition. Perceptions are based upon one's senses and are typically correct. Intuitions, on the other hand, are "mental representations, ideas, hypothetical solutions [that] may be biased, distorted, incomplete, vague or totally wrong" (Fischbein, 1997, p. 28). Thus, it is possible to differentiate *perception of* a situation from *intuition about* the situation. Connecting back to the context of a board game, perceptions of a situation include the location of a player's pieces on the board, the amount of resources that other players have acquired, and the number of victory points one has acquired.

Intuitions, on the other hand, might include the likelihood that a player's current strategy will be successful, the anticipated moves of an opponent, or the chances that the outcome of a random process will be beneficial.

Another distinction made by Fischbein was between an intuition and a skill. An individual can be highly skilled at something without having any particular intuitions for that activity. For example, one can be very skillful at calculating probabilities for the sum of two six-sided dice without having any intuitive ideas about the roll. Intuition, on the other hand, "*is more than a system of automatized reactions, more than a skill or system of skills; it is a theory, it is a system of beliefs, of apparently autonomous expectations"* (Fischbein, 1987, p. 88). To this end, experience plays a critical role in shaping intuitions because it has the potential to stabilize expectations. Fischbein noted that expectations could become "so stable, so firmly attached to certain circumstances, that their empirical origin may, apparently, vanish from the subject's awareness" (p. 88). Whereas skills are learned through intentional practice, intuitions can be learned unintentionally through repeated experience. Consequently, the origin of an intuition may be unclear to the person holding it.

In his 1987 book, *Intuition in Science and Mathematics,* Fischbein reviewed the presence of intuition in a variety of literature including mathematics, science, philosophy, and art. He noted that the definition of intuition was varied across different disciplines and included descriptions of artistic clairvoyance, religious revelations, and scientific discovery. Drawing across all of these different examples, Fischbein identified nine properties of intuition that were shared across contexts. The first of these properties is that intuitive knowledge is immediate and self-evident. A summary of the properties is provided in Table 1.

Property	**Description**
Self-evidence and Immediacy	An intuitive cognition appears subjectively to the individual as directly acceptable. The individual does not see the need for extrinsic justification either in the form of a formal proof or empirical support.
Intrinsic Certainty	Intuitive cognitions are accepted as certain by the individual. Self-evidence and certainty are highly correlated but they are not the same thing. Certainty does not imply self-evidence nor does self-evidence imply certainty.
Perseverance	Intuitive cognitions are robust. Formal instruction aimed at providing conceptual knowledge can have little impact on an individual's intuitive background knowledge. It is possible for an individual to simultaneously hold erroneous intuitions and conceptual interpretations.
Coerciveness	Intuitions strongly affect an individual's reasoning by appearing to be absolute, unique representations or interpretations. Alternative interpretations are typically excluded or resisted.
Theory Status	An intuition is held by the individual as a theory and expressed in a particular representation using a model (paradigm, analogy, diagram, etc.). It is not a skill or perception.

Extrapolativeness	An intuition exceeds the available data. It is an extrapolation beyond the information at hand. Extrapolation alone is not enough to define an intuition, there must also be a feeling of certainty.
Globality	An intuition offers a global, synthetic view to the individual. It is concerned with the whole not the parts.
Implicitness	The activity of intuition is generally unconscious and the individual is only aware of the final product (i.e., the self-evident, intrinsically consistent cognitions).
Cognitive-Behavioral Function	Intuitions have the same behavioral function as perceptions; however, they are at a symbolic level. Consequently, intuitive cognitions prepare and guide both mental and practical activity.

Table 1. Nine Properties of Intuitions (Adapted from Fischbein, 1987)

In addition to these general properties, Fischbein also created two classification systems for intuitions. The first system categorizes intuitions by considering the relationship between an intuition and a solution to a problem. Using this scheme, an intuition falls into one of four categories: affirmatory, conceptual, anticipatory, or conclusive. The second classification system considers whether or not a particular intuition developed within the context of systematic instruction. This article focuses on Fischbein's second categorization (a more detailed description of the first categorization can be found in Fischbein, 1987).

Fischbein's second categorization system classifies intuitions developed independently of any systematic instruction as primary intuitions. These intuitions, he explained, were a result of one's personal experience.

> Our term 'primary intuitions' does not imply that these intuitions are innate, or *a priori.* Intuitions, both primary and secondary, are in fact learned cognitive capacities in the sense that they are always the product of an ample and lasting practice in some field of activity. (Fischbein, 1987, p. 69)

Thus, primary intuitions are those that individuals develop in out-of-school experiences. As noted, in order for these intuitions to develop there must be "ample and lasting practice." Consequently, primary intuitions should not be regarded as developing from one encounter with a particular idea, but rather as developing over the course of many such encounters.

Fischbein described secondary intuitions as cognitive beliefs that resulted from systematic instruction. This instruction must actively involve the learner in order for an intuition to develop. He noted,

> Such a process implies, in our view, the personal involvement of the learner in an activity in which the respective cognition play the role of necessary, anticipatory and, afterwards, confirmed representations. One may learn about irrational numbers without getting a deep intuitive insight of what the concept of irrational number represents. Only through a practical activity of measuring may one discover the meaning of incommensurability and the role and meaning of irrational numbers (Fischbein, 1987, p. 202)

As is evident in the passage, Fischbein stressed that secondary intuitions are formed only when an individual is involved in an activity that requires serious consideration of a particular idea. Thus, instruction that fails to actively engage the learner in meaningful analysis of a particular concept will not result in the development of secondary intuitions. Following from Fischbein's classification, both primary and secondary intuitions are *learned* through experience and develop throughout one's lifetime.

Dual-Process Theory

Dual-process theory has its roots in cognitive psychology (Kahneman, 2002) and has only recently been applied to mathematics education (Leron & Hazzan, 2006, 2009). The theory, which has grown from Amos Tversky and Daniel Kahneman's (1974) work concerning heuristics and biases, characterizes the mind as having two distinct systems, System 1 (S1) and System 2 (S2).

Like Fischbein's theory, dual-process theory concerns intuitions, but it takes a different perspective on their role in cognition. Dual-process theory considers the relationship between intuitive (immediate) and analytical modes of thinking and behavior. The central principle of dual-process theory is that cognition and behavior "operate in parallel in two quite different modes...roughly corresponding to our common sense notions of intuitive and analytical thinking. These modes operate in different ways, are activated by different parts of the brain, and have different evolutionary origins" (Leron & Hazzan, 2006, p. 108). The main difference between S1 and S2 is related to their accessibility. S1 is regarded as being "halfway between perception and (analytical) cognition" (p. 108). Its processes are fast, automatic, effortless, unconscious, and difficult to change. Additionally, S1 tends to contextualize and personalize problems. Decisions made by S1 are closely tied to the context of a problem.

S2 processes, on the other hand, are slow, conscious, flexible, and require effort. Unlike S1, S2 removes context and depersonalizes problems. It is more capable of creating rule-based representations and identifying underlying principles. Additionally, S2 can consider problems outside of a context.

The two systems are not isolated from each other. It is possible for skills, such as playing a particular board game, to migrate from S2 to S1. Initially, playing a board game requires a great deal of effort for novices because they must attend to the rules, understand how in-game actions are influencing gameplay, etc. However, as individuals repeatedly play a game and internalize the boundaries of the system, this skill may migrate from S2 to S1. As the context becomes familiar, in-game decisions, which were initially complicated and required a great deal of consideration, can transition from S2 to S1. Likewise it is possible for skills to migrate from S1 to S2. For example, walking along a straight line is normally handled by S1 in adults. However, if an adult is very tired this skill can require a great deal of effort and transition from S1 to S2.

Although the two systems are viewed independently, they often work together. S1 provides quick, automatic responses to appropriate situations while S2 serves as a monitor and critic of S1. However, this coordination between the two systems does not always operate well. S2 requires more effort and energy than S1. This means that, from a conservation of resources standpoint, S2 should only be used when there is a clear need. Research has documented problems in which S2 fails to engage for the majority of people (Kahneman, 2002; Leron & Hazzan, 2006). In these problems, an issue arises when S1 produces a quick (and often incorrect) response and S2 does not serve as an effective monitor. The Bat Problem, reported by Kahneman (2002) is one example,

> A baseball bat and ball cost together one dollar and 10 cents. The bat costs one dollar more than the ball. How much does the ball cost?
>
> Almost everyone reports an initial tendency to answer '10 cents' because the sum \$1.10 separates naturally into \$1 and 10 cents, and 10 cents is about the right magnitude. Frederick found that many intelligent people yield to this immediate impulse: 50% (47/93) of Princeton students, and

56% (164/293) of students at the University of Michigan gave the wrong answer. (p. 451)

According to dual-process theory, the specifics of this problem (i.e., the total cost of $1.10 and the bat costing $1 more than the ball) cause it to be solved quickly and erroneously by S1. The context appears straightforward and, for many people, S2 is not critical of the answer supplied by S1 because the problem does not present itself as needing such oversight. Thus, most individuals provide an incorrect answer of ten cents. Those who answer the problem correctly, in contrast, likely do so because of the involvement of S2 as either a critic of S1 or as the primary reasoning system. Research has shown that increasing the difficulty of the problem by changing the numbers yields more correct responses. It has been argued that this change in difficulty triggers the involvement of S2. Leron and Hazzan (2006) noted that many of the problems explored in probability research can be explained in terms of dual-process theory. They suggest that non-normative (i.e., different from an expert) responses to some problems may be the result of a failure by S2 to monitor S1.

The Integrated Perspective

Although the theory of intuition and dual-process theory are distinct, they complement each other. Fischbein's theory describes the characteristics of intuitions as well as their formation; however, it does not provide a detailed description of how these intuitions impact decisions. Dual-process theory, on the other hand, is concerned with the decision-making process rather than the specifics of components. These two theories can be combined to form an integrated perspective, which is modeled in Figure 2.

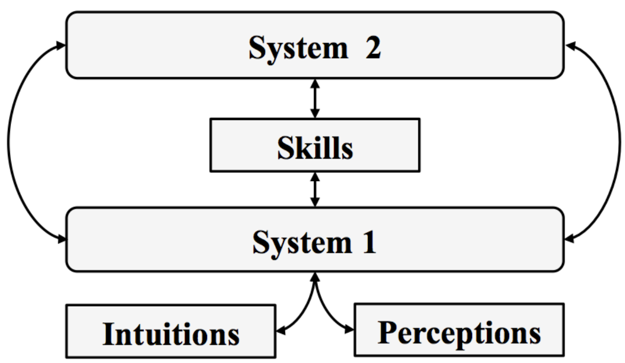


Figure 2. The integrated perspective

In this model, S2 is the higher order reasoning system with the capability to generalize and reason abstractly. However, this comes at a cost of effort and energy. S2 also serves as a monitor of S1. S1 is the immediate, context-based decision system. It's decisions come at a lower cost in terms of energy and effort but can be misguided by intuitions, perceptions, or skills. Intuitions are regarded as cognitions that result from either experience or systematic instruction and have the nine properties outlined by Fischbein (see Table 1). They are immediately available to S1 in a given situation. Perceptions are cognitions that are formed based upon available sensory information and immediately available to S1. Both intuitions and perceptions may be considered by S2, but they do not prompt direct action. Skills are effortful actions that are initially controlled by S2. If a skill is sufficiently practiced, it can become an automatic, internalized reaction that is available to S1.

There are a number of reasons this perspective is appropriate for use in studying an individual's thinking as they play board games. First, the perspective allows for the examination of immediate reactions (controlled by S1) as well as long-term strategies (controlled by S2). For example, in *Settlers of Catan* a roll of 7 triggers the movement of the robber. The player who rolled the 7 must move the robber to a new tile on the board and is then able to take resources from an opponent that controls a building on the tile. For some players, this action of moving the robber is immediate and based upon intuitions and perceptions (S1). These players focus on immediate risks or rewards. However, for others moving the robber is a much more analytical process (S2) that draws on information from a variety of sources (e.g., current game conditions, history with opponents). These players consider how moving the robber will influence their overall strategy. Consequently, their actions may defer immediate rewards in order to increase the likelihood of long-term success. Such a player may choose to use the robber as a tool to ally himself or herself with another player even though this may have negative short-term consequences.

Second, the integrated perspective provides an avenue for examining how the intuitive ideas that individuals hold about probabilistic phenomena as well as their perceptions influence in-game decisions. For example, an intuitive understanding of the probabilities of different six-sided dice roles may be an advantage in a game that incorporates quick decisions about dice outcomes (e.g., selecting tiles that are more likely to be activated). Perceptions may reinforce this intuition or appear to contradict it. For example, if a player observes a series of sums totaling two, they may ignore their intuitive understanding about randomness and adopt a misconception (e.g., the dice are not behaving normally).

Third, the perspective allows for opportunities to observe the migration of skills from S2 to S1. As noted, board game RL situations initially require a great deal of effort to understand because players must coordinate allowable actions with the likelihood of success. Consequently, these decisions are made by S2. However, as individuals repeatedly play a particular board game the in-game situations become familiar. The allowable actions become internalized and the players shift from novices towards experts. Thus, these repeated experiences with probabilistic phenomena may change from analytic cognitions made by S2 to intuitive cognitions made by S1.

Finally, the perspective provides a lens on the formation of primary intuitions. Playing a game requires individuals to repeatedly anticipate the likelihood of events and then react to the outcome. These decisions are not optional but a necessary part of gameplay. Moreover, variations of a particular event are typically repeated many times over the course of a game. Thus, players may begin to develop primary intuitions as the result of repeated play. These intuitions are considered primary because the experiences with board games are not within a formalized instructional setting.

Vignettes for Illustration

The following vignettes have been created to illustrate how this integrated perspective might be used as a lens to examine reasoning about risk within a board game setting. Although the vignettes are fictitious, they have been created to model actions and conversations that typically occur in playing the game *Settlers of Catan*.

Vignette 1: A Discussion During Game Setup

Erin, Kai, and Henry have decided to play *Settlers of Catan*. Erin and Kai are very familiar with the game having played it frequently over the last several months. Henry has only played the game once.

> *Henry:* I think I'm going to start with one town at the three-for-one port that is at the vertex of the fields and pasture tiles. I will be able to collect a lot of one resource and trade for another resource more easily.
>
> *Kai:* It's not worth your time to build on that port.
>
> *Henry*: Why do you say that?

Kai: Because you can only get resources from two tiles instead of the usual three. I don't think it is worth the risk just to gain the three-for-one trade ability.

Henry: You don't know that the town won't produce. The numbers, 9 and 10, aren't terrible.

Erin: Yeah, just because the tiles aren't a 6 or 8 doesn't mean they won't activate. I've seen a lot of games where 6 and 8 are hardly ever rolled.

Kai: I'm not saying that it won't produce. I'm saying that three-for-one isn't a really great advantage when you have to limit yourself to two tiles. You can trade four-for-one with the bank at any time or get a better exchange with other players.

The vignette begins with Henry articulating his strategy. He is attempting to develop a generalized strategy, which is focused on having a better trade ratio, that might benefit him throughout the game. Kai points out a risk of the strategy, i.e., building on the port means the town will only be on two tiles. Henry and Erin respond to Kai's comments with a focus on the likelihood of rolling the numbers on the tiles. This allows some information concerning their primary intuitions about probability. Henry's response that "the numbers, 9 and 10, aren't terrible" suggests he knows that 9 and 10 are less likely than sums of 6 or 8, but more likely than a sum of 2 or 12. Thus, although his original statement appeared to be focused on a trade deal, there is some evidence that he also considered the likelihood of the tiles producing resources. Erin's response appears to outline a primary intuition about the frequency of sums in a game that she has learned through repeated play. A question that might be asked is whether her response shows this experience has negatively influenced her probabilistic reasoning by suggesting that theoretical probability is not connected to real-world practice or positively influenced her reasoning by reinforcing that individual outcomes of a random event are uncertain. In his follow-up comments Kai indicates that he believes the risk prompted by not having a third tile is not worth the reward of a three-for-one trade ability. His focus is on the likelihood of a town generating resources when it is on three tiles versus when it is on two tiles. Overall, this excerpt demonstrates S2 at work in trying to generalize and abstract a RL situation. The comments reveal how each player's perceptions of the board and probabilistic intuitions are being used by S2 to form a general strategy.

Vignette 2: A Midgame Discussion

As play progresses, Henry develops an early lead with seven points. Erin is close behind with six points (recall that the first player to score 10 points is the winner). On his turn, Kai rolls the dice and gets a sum of seven. This triggers movement of the robber. He decides to move the robber onto a tile that is bordered by towns from both Henry and Erin. As part of the robber action, Kai must take a card from one of the players and he decides to take a card from Erin.

Erin: Why are you taking a card from me? Henry is in the lead.

Kai: I don't think he is going to win.

Erin: He's only four points from victory!

Henry: And I've been able to use my port effectively to get goods.

Kai: Your towns are mostly on tiles with unlikely numbers. So I doubt you're going to keep benefiting from dice rolls. Besides you need to upgrade some of your towns to cities in order to win and even if your tiles produce they wouldn't provide the right resources.

Henry: Well I disagree. The dice have been in my favor today.

Erin: It's like I said before, you really can't predict the roll. In some games a 6 and 8 are hardly ever rolled. That's just how it goes.

Henry: Yeah, I've been doing a pretty good job of rolling what I need. It's all technique.

Kai: (Shaking his head). Look, Erin is just about to claim the longest road card giving her two more points. She is also close to having the largest army card, which is another two points. She can end this game in two rounds.

Henry: We'll see about that. (Rolls a seven). Ha, see that? Told you I've got the skills. (He quickly moves the robber to a tile bordered by Kai and takes a card).

Kai: Why would you do that? Erin is clearly in the lead here.

Henry: I disagree. Besides you have a lot of cards.

Kai: But only one of my cards is a resource that you need for upgrading. Most of Erin's cards are resources that you need. That move doesn't make sense.

This vignette demonstrates probabilistic decisions being made *in-action* by players. Kai's comments indicate that he is assessing the overall progress of the game and weighing the states of each opponent (S2). His assessment that Henry is not likely to win is based on information about the likelihood that Henry's towns will produce combined with knowledge that even if the towns did produce the resources would not be immediately helpful. Henry's comments indicate that he is focused on more immediate conditions (S1). He also appears to have intuitive beliefs about controlling the outcome of a dice role indicating probabilistic misconceptions. Moreover, his reaction to the roll of a 7, which focuses on taking a card from Kai, is a quick judgment that appears to be in response to Kai's comments rather than based on in-game conditions. For her part, Erin's comments suggest that she is aware that the individual outcome of a random event is uncertain. Overall, this excerpt demonstrates how players might be using S1 and S2 to make decisions about a given situation. Kai's decisions indicate that he is still trying to generalize and predict future outcomes. Henry, on the other hand, appears to be making quick, in-the-moment decisions without focusing on the overall risks presented by Erin.

Concluding Thoughts

This article has sought to illuminate one possible avenue for studying individual's thinking about risk. As with any research agenda, there are some initial challenges that must be addressed. In particular, researchers must decide which individuals to include, which board games to use, and the specific RL situations they will investigate. For example, it seems likely that such investigations would occur both within and outside of traditional classrooms and draw on a small set of games with a limited range of fortuity and intricacy. Additionally, it is probable this research would involve individuals from a variety of age ranges some of whom are likely well passed school age. The study of such environments will require frameworks that have the flexibility to examine how various cognitions (intuitions, perceptions, skills, etc.) influence both long-term and short-term decisions. The integrated theoretical perspective outlined previously may offer this flexibility; however, it should be seen as an initial framework in need of refinement and extension.

Challenges not withstanding, I believe that the possibilities of this medium are exciting. Work within the field of mathematics education has already laid the groundwork for meaningful investigations and the popularity of board games has grown greatly in the last decade. If the field wishes to investigate thinking about this topic, it seems reasonable that we do so in ways that can model the prevalence of probability in modern life and the diversity of contexts in which individuals encounter RL situations. Moreover, the experiences should have some authenticity for participants and be dynamic. Board games offer one potential avenue with a great diversity in design and components. This medium has the potential to engage individuals in authentic RL situations while at the same time providing researchers with some control over the system. Thus, the atmosphere appears right for mathematics education researchers to take advantage of the medium of board games for studying RL situations. As a community, it is time to take a seat at the table, roll the dice, and make the next move.

References

Ayton, P. & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory and Cognition, 32*(8), 1369–1378.

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*, 113–123.

Chernoff, E. J, & Sriraman, B. (Eds.) (2014). *Probabilistic thinking: Presenting plural perspectives*. Berlin, Germany: Springer Science.

Cuzick, J., Thorat, M. A., Bosetti, C., Brown, P. H., Burn, J., Cook, N. R., …, Umar, A. (2014). Estimates of benefits and harms of prophylactic use of aspirin in the general population. *Annals of Oncology*. Advance online publication. doi:10.1093/annonc/mdu225

Dawes, R. M. (1988). *Rational choice in an uncertain world*. New York, NY: Harcourt, Barce, and Jovanovich, Inc.

Fischbein, E. (1987). *Intuition in science and mathematics.* Boston, MA: D. Reidel Publishing Company.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions?: An exploratory research study. *Educational Studies in Mathematics, 15*, 1–24.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics, 22*, 523–549.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*, pp. 96–105.

Gold, E. (1997). *The gambler's fallacy.* (Unpublished doctoral dissertation). Carnegie Mellon University, Pennsylvania.

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. Retrieved from http://www.nobel.se/economics/laureates/2002/kahnemann-lecture.pdf

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*, 392–414.

Leron, U., & Hazzan, O. (2006). The rationality debate: Application of cognitive psychology to mathematics education. *Educational Studies in Mathematics*, *62*, 105–126. doi:10.1007/s10649-006-4833-1

Leron, U., & Hazzan, O. (2009). Intuitive vs analytic thinking: Four perspectives. *Educational Studies in Mathematics*, *71*, 263–278. doi:10.1007/s10649-008-9175-8

Metz, K. E. (1997). Dimensions in the assessment of students' understanding and application of chance. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 223–238). Amsterdam, The Netherlands: IOS Press.

Mooney, E. S., Langrall, C. W., & Hertel, J. T. (2014). A practitional perspective on probabilistic models and frameworks. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 495–508). Berlin, Germany: Springer Science.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York, NY: Macmillan.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.

Uman, M. A. (1986). *All about lightning.* New York, NY: Dover Publications.

Developing Strategic and Mathematical Thinking via Game Play: Programming to Investigate a Risky Strategy for Quarto

Peter Rowlett
Nottingham Trent University, U.K.

Abstract: The Maths Arcade is an extracurricular club for undergraduate students to play and analyse strategy board games, aimed at building a mathematical community of staff and students as well as improving strategic and mathematical thinking. This educational initiative, used at several universities in the U.K., will be described. Quarto is an impartial game played at the Maths Arcade, in that there is one set of common pieces used by both players, and one where stalemates are a common outcome. While some students play without apparent direction until a winning opportunity appears, others adopt a more risky strategy of building the board towards a winning position, which could allow either player to win. Whether building towards a win is a sensible strategy, when the other player could equally well benefit, is a topic of debate at the Maths Arcade. Intending to suggest a possible student project, this article will describe a method to represent Quarto as an array of binary numbers, making the game suitable for programming in Python. Then, one strategy is programmed to play at random unless a winning move becomes available, while another is programmed to work towards a winning position. These are calibrated by playing against a completely random strategy and against themselves, then they are played against each other. The more risky strategy is found to win over the more naive player in around two thirds of one million games. Some limitations and possible areas of development are discussed.

Keywords: mathematical thinking, strategy, risk, games, programming.

Introduction

The Maths Arcade is an extra-curricular activity to involve students and staff in playing and analysing strategy games. A game played at the Maths Arcade, Quarto, is described, along with playing strategies. One strategy is identified as more risky, because it involves working to move the board to a position where either player could win. Intending to suggest a possible student project, Quarto is represented in a way suitable for programming and this is used to play the risky strategy against a less risky behaviour.

An attempt is made to give outline information such as might be useful to a student embarking on project work in this area (and, indeed, investigations have not gone much further). In this way, it is hoped that this article might form the basis of such project work.

The Maths Arcade

The Maths Arcade was initiated by Bradshaw (2011) as a weekly, extra-curricular drop-in session where students and staff play a variety of strategy games and puzzles. This aimed particularly to support students who are new to university by providing an environment where they could interact with other students and staff as part of a mathematical community. Having observed some students reluctant to attend ‘help sessions’ due to a perception that these are for weaker students, Bradshaw designed the Maths Arcade to both support weaker students and stretch more confident learners. As such, the Maths Arcade is an opportunity to develop mathematical thinking and problem-solving skills in a situation

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 55-61

where prior mathematical knowledge is not particularly relied upon. Different approaches to running the Maths Arcade at various universities in the U.K. are explored by Bradshaw and Rowlett (2012).

At Nottingham Trent University, a description of setting up the Maths Arcade is given by Rowlett and Webster (2013). Students' experience of the Maths Arcade begins as a 'getting to know each other' induction activity for mathematics undergraduates in the first week of term. This continues as a weekly drop-in session in term time during the three years of the degree. Games might be simply played, as a fun activity in a mathematics-themed social support environment, while some students start to think about strategies. Questions which arise in game play include: What is the best strategy? Is there an advantage to a particular position? Is there a benefit to going first? As well as a useful induction activity and an extra-curricular curiosity, at Nottingham Trent University students undertake a substantial individual project in the final year of their degree, and some have chosen to study game theory through attempted analysis of a Maths Arcade game.

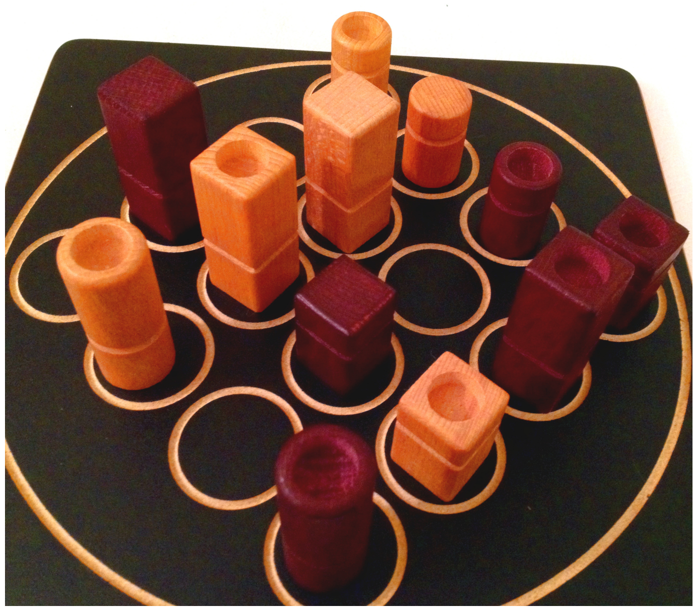


Figure 1. A Quarto board and some pieces.

Quarto

Game play

Quarto is an impartial game of perfect information played at the Maths Arcade. This means there is one set of common pieces used by both players, and each player is perfectly informed of all the events that have previously occurred in the game at all times.

Quarto is played on a 4 × 4 board. Each game piece has four attributes each taking one of two values. Specifically, it is: white or black; short or tall; round or square; and, having a dimpled top or flat. Each combination is used, so the game uses $2^4 = 16$ pieces. This means the number of pieces matches the number of spaces on the board. A Quarto board is shown in Figure 1.

Game play is that players take turns to choose an unplayed piece for their opponent to place in an unused space on the board. Thus, there are two stages to each turn, the same for both players: 1) play the piece handed over by the opponent; 2) hand a piece to the opponent.

The aim is to be the player who places the fourth piece in a line (row, column or diagonal) which all match in any one attribute (i.e., four that are square, or four that are dimpled, etc.). A stalemate is possible, in which case the board is filled with no winning lines and the game is a draw.

Strategies

A reasonable strategy for a new player is outlined below. Call this strategy 'naive'.

1. Play the piece handed over by the opponent:

    (a) play a winning position if handed a winning piece;

    (b) otherwise, play randomly.

2. Hand a piece to the opponent:

    (a) avoid handing over a winning piece for your opponent to play;

    (b) otherwise, choose randomly.

As a more mature strategy, some attendees at the Maths Arcade try to build lines of like pieces. That is, pieces which match in at least one attribute. This strategy is outlined below. Call this strategy 'risky'.

1. Play the piece handed over by the opponent:

    (a) play a winning position if handed a winning piece;

    (b) otherwise, play to build a line of like pieces if possible;

    (c) otherwise, play randomly.

2. Hand a piece to the opponent:

    (a) avoid handing over a winning piece for your opponent to play;

    (b) otherwise, choose randomly.

The risky strategy attempts to increase the options for winning moves, thus attempting to direct the board to a winning position. This is more risky because building lines that could win advantages both players. However, even though it is an advantage to both, players playing this strategy appear to win more often than those following the naive strategy – excluding human error. (Students at the Maths Arcade are encouraged to 'rewind' silly mistakes and re-play, in order to learn and study strategy.) Since stalemates are common among experienced players, whether this strategy holds an advantage is a topic of debate at the Maths Arcade.

For the risky player, if a winning position cannot be taken then an attempt is made to match like pieces. One strategy for doing this is outlined below.

1. Search the board for a line where the piece to be played has at least one attribute value in common with those already on that line.

2. If such a position is found, check that playing there would not create a situation where there exists a no-win scenario for that attribute. That is, a situation where one line of three pieces match in one value of an attribute and, simultaneously, another line of three pieces match in the other value for that attribute. If one line would win with a black piece and another line would win with white, this results in a loss since every piece is either black or white.

	1101	1111	
	0101	0011	0010
	0000		

Figure 2. A Quarto game in progress, represented by binary strings.

The choice made here is to search as described until the first opportunity to play a like piece arises, then play it.

For testing purposes, a third strategy is proposed: random play. A player simply chooses an available board position at random for each piece handed over, and chooses a piece to hand over at random from the remaining pieces. This is extremely poor strategy and will result in the player not necessarily winning even when handed a winning piece, and handing over a winning piece even when non-winning alternatives are available.

Programming Quarto

Quarto can be represented using four-character binary strings. First, label each attribute value either 0 or 1: white (0) or black (1); short (0) or tall (1); round (0) or square (1); and, dimpled (0) or flat-topped (1). Using a positional system in this order, attributes for a piece can be represented. For example, 0100 would be white, tall, round and dimpled; 1001 would be black, short, round and flat-topped.

Then a winning condition is met if four pieces in a line share at least one digit in common. For example, 0100 and 1001 are both round, so both have the third digit 0.

A board position can be represented by a 4×4 grid of such binary strings (an example is shown in figure 2). In figure 2, if the next player plays piece 0001 into the 3rd row, 1st column position, then a line of four will be created which all match in the leading digit, 0. Such a grid could be stored in an array or list.

If a game system keeps track of the state of the current board and the pieces available to play, then players can query these as necessary, since each player in Quarto operates from a position of perfect information. Then the random player can simply select at random from the available pieces or spaces (depending on the phase of the game), while the naive and risky players can search through the board and the unplayed pieces as necessary to meet their strategic objectives.

For example, both the naive and risky players operate on the principle of playing in a winning board position if handed a winning piece. As a brute force approach, they could check each line which contains three pieces against the piece to be played. If a set of four pieces is thus discovered which matches in one digit, then the piece is played and the game won. If no such set is found, the piece may be played according to the next move in the strategy.

Playing to simply avoid handing over a winning piece would operate on a similar principle of matching each available piece against lines of three on the board until a piece is found that does not match any line, if this is possible.

Method

Three players (random, naive and risky) programmed in Python were run in a series of trials.

Trials took the form of 10^6 games of Quarto with player 1 (P1) playing against player 2 (P2). This was arranged so that P1 and P2 take turns to go first, to negate any first player advantage or other imbalance.

A large number of games were run because all strategies have a random element, in choice of piece to hand over and in where to play. 10^6 was chosen as a suitably large number that would compute in a reasonable amount of time (such that the whole process took hours, not days, to complete; though certainly the program could be made more efficient).

First, to check the system, three trials were run in which both players took the same strategy. As these games are symmetric (same strategy, equal times playing first), we expect roughly equal numbers of wins for each player.

Then, to confirm that the naive and risky strategies are sufficiently better than chance, trials were run with each against a random player. As the random player will not necessarily win even if handed a winning piece, and may choose to hand over a winning piece when alternatives are available, we expect the more elaborate strategies to win much more often.

Finally, the risky strategy is played against the naive. As the risky strategy is building lines for both players to take advantage of, and otherwise both adopt similar behaviour, we might expect roughly equal results.

Results

The results from the three trials with both players taking the same strategy are given in Table 1. In each case, players won roughly the same number of games, with a small number of stalemates.

P1	P2	P1 wins	P2 wins	Stalemates
random	random	490668 (49.07%)	489600 (48.96%)	19732 (1.97%)
naive	naive	495707 (49.57%)	496734 (49.67%)	7559 (0.76%)
risky	risky	498650 (49.87%)	499778 (49.98%)	1572 (0.16%)

Table 1. Results from each strategy played against itself 10^6 times.

The results from the two trials with P2 taking a random strategy and P1 taking naive and risky strategies are shown in Table 2. Both more elaborate strategies win substantially more games than the random player, though the risky player wins more often and loses less often than the naive player did, resulting in fewer stalemates.

P1	P2	P1 wins	P2 wins	Stalemates
naive	random	974407 (97.44%)	23872 (2.39%)	1721 (0.17%)
risky	random	993154 (99.32%)	6356 (0.64%)	490 (0.05%)

Table 2. Results from running risky and naive strategies against random play, each pair 10^6 times.

The results from playing the risky strategy against the naive, shown in Table 3, are that the risky player won about twice as often as the naive.

P1	P2	P1 wins	P2 wins	Stalemates
risky	naive	678113 (67.81%)	317313 (31.73%)	4574 (0.46%)

Table 3. Results from running risky strategy against naive strategy 10^6 times.

Discussion

The Maths Arcade is an extra-curricular activity used at Nottingham Trent University and several other U.K. universities to involve students and staff in playing and analysing strategy games. A program was written in Python to simulate play of Quarto, a game used at the Maths Arcade. Inspired by game play observed at the Maths Arcade, two strategies were programmed and tested against each other. The first played naively, winning if possible but otherwise proceeding randomly. The second actively tried to move the board into a winning position. As Quarto is an impartial game of perfect information, the second strategy invites either player to win, so was described as more risky. When these strategies were tried against each other 10^6 times, the more risky player won around twice as many games as did the naive player.

It is not possible to conclude that the more risky approach to Quarto is more likely to result in a win, since the result could have been affected by how well the strategies were programmed. Certainly, the strategy employed to match like pieces was not very sophisticated and could be improved. The choice here was to search until the first opportunity to play a like piece arises, then play it.

This strategy is sensitive to the order in which the board is searched and may cause the player to block a better option to match a weaker one. For example, in Figure 3 the player creates a column of two like pieces despite the fact that doing so blocks a row of three like pieces. If it had played in another position in the same column, it would have retained both options.

			1110
0000	0001	0010	<u>1100</u>

Figure 3. The risky algorithm as programmed may play the underlined piece as indicated in order to make a column of two with 1110, even though it blocks a row of three matching in two attributes 00**.

The strategy outlined here does not search ahead for which pieces may be handed to the opponent, and this will cause losses. For example, it is a strategic error to make a line of three black pieces when every piece left to be played is also black.

The strategy outlined also does not take into account other lines of three on the board, and this will cause losses too. For example, it is a strategic error to make a line of three black pieces when there is already a line of three square pieces on the board and every piece left to be played is either black or square.

Other considerations have impact on the strategy, and there are alternative strategies that could be analysed. For example, one player at the Nottingham Trent University Maths Arcade tends to play to block potential winning lines, reducing the number of possibilities rather than working to increase them, with some successes.

Another interesting avenue of investigation is the number of stalemates that arise. We might expect that, as players get better, stalemates may become more common (in parallel with Tic Tac Toe), and this is observed at the Maths Arcade, but this is not apparent from the data here. For example, it is interesting that the naive player against random finds more stalemate positions than risky against random. And, indeed, that random playing against random finds more stalemates than the other strategies playing themselves. The cause may be that the risky player is building opportunities to win, but not necessarily being very careful about whether the win may occur on the opponent's move. A more sophisticated pair of players might well generate more draws.

A student wishing to undertake a project in this area would be well advised to investigate relevant areas of combinatorial game theory. Various claims are made in this article that should be identified and verified. Where strategy decisions have been described, these should be re-considered. Work might begin by developing a game system capable of keeping track of the board and available pieces, then running two players with random play, before attempting to program more sophisticated strategies. Such a game system should be designed in a way so that different game strategies can easily be assigned to each player. It should also have routines so that a single game can be displayed in progress graphically or multiple games can be played with only a summary of wins, losses and draws at the end displayed. The former option would be to observe game play is as expected, while the latter is for simulation work.

Acknowledgements

The Maths Arcade at Nottingham Trent University was supported by a grant from the Institute of Mathematics and its Applications.

References

Bradshaw, N. (2011). The University of Greenwich Maths Arcade. *MSOR Connections, 11*(3), 26-29. Retrieved: journals.heacademy.ac.uk/doi/abs/10.11120/msor.2011.11030026.

Bradshaw, N., Rowlett, P. (2012). *Maths Arcade: stretching and supporting mathematical thinking*. Birmingham, U.K.: Maths, Stats and OR Network. Retrieved: www.mathcentre.ac.uk/resources/uploaded/mathsarcade.pdf.

Rowlett, P., & Webster, K. (2013). A Maths Arcade at Nottingham Trent University. *Mathematics Today 49*(3), 120.

Risk Education
A Worldview Analysis of What is Present and Could Be

Gale L. Russell

University of Regina, Canada

Abstract: Risk, risk analysis, risk management and risk-based decision-making are ubiquitous ideas in the modern world. Consequently, risk education is emerging as a new field of research. However, just as the defining of risk and what it entails is a contested topic, so too is the field of risk education research open to many possible approaches. In this paper notions of risk, particularly as they play out in research on risk education, are analyzed (within an ethical space) using a theoretical framework based on the Traditional Western and an Indigenous worldview. Through this analysis, along with the identification of the kinds of knowledge and ways of knowing currently being valued in the research, other kinds of knowledge and ways of knowing that may prove just as important emerge.

Keywords: decision-making, ethical space, indigenous, risk, risk Education, western, worldview.

Introduction

> No matter how dominant a worldview is, there are always other ways of interpreting the world. Different ways of interpreting the world are manifest through different cultures, which are often in opposition to one another. One of the problems with colonialism is that it tries to maintain a singular social order by means of force and law, suppressing the diversity of human worldviews (Little Bear, 2000 p. 77).

Risk, risk analysis, risk management and risk-based decision-making are ubiquitous ideas and terms in the modern world; however, what is meant by risk and how one can analyze, manage, or make decisions based upon risk remains, if not contended, then inconclusive. Borovcnik and Kapadia (2011), define risk as "a situation with inherent uncertainty about the (future) outcomes, which are related to impact (cost, damage, benefit)" (p. 5503), a definition which will work for the purpose of this paper. Within much of risk research, there is general agreement that "there are two levels of criteria for making decisions: personally preferred ones and rationally bound ones" (p. 5503). There even is agreement as to the origins of the knowledge held in relation to each of the criteria (affective/emotional responses and scientific methods, respectively); however, the perceived worth of each of these criteria by researchers is not so clear cut. Further, as with other emerging and prominent features of the modern world, like technology, there is a proclaimed need for the study of risk to be part of students' educational experiences. Hence, the need to consider risk education: where does it belong, what should it look like, and how can it happen – researchers have begun to investigate this proposition.

Concurrently, within another emerging area of research, it has been theorized that the grounding of the teaching and learning of mathematics within an Indigenous worldview (the transreform approach) has the potential to bring about changes that grounding within the Traditional Western worldview might not (Russell & Chernoff, 2011, 2012, 2013a, 2013b). Particularly, it has been proposed that such a change in worldview could result in the end of the Math Wars (Russell & Chernoff, 2013b); the

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 62-84

acceptance and implementation of research-supported approaches to the teaching and learning of mathematics, such as those related to constructivist learning theory (Russell & Chernoff, 2012); and to diminish, even eliminate, the marginalization of students (Indigenous and non-Indigenous alike) with respect to mathematics (Russell & Chernoff, 2012, 2013b). The goal of this paper is to provide an analysis as well as examples, which illustrate the importance of extending the current boundaries of risk education research to include considerations that emerge, yet untapped, through the analysis. The purpose of this paper is to use this same theoretical framework to analyze the existing risk education research in order to provide new insights into what has been considered and what has not. Further, through the use of particular examples of situations involving risk, an argument is made for how risk education could be strengthened in relation to this analysis.

This paper begins with a brief discussion of key areas and ideas related to risk and risk education. Then, the aforementioned theoretical framework and an explanation of how it will be applied are described. From that point, the theoretical framework is used to analyze current research in risk education, highlighting what kinds of knowledge and ways of knowing are being valued, which are not being valued, and which remain hidden. Finally, two examples of recent incidents and the risk knowledges that were valued and not valued within them are presented to demonstrate the importance, the possibilities, and the consequences of the analysis results.

Risk in the Literature

Existing research and theories about risk are abundant and variable. Within different fields (psychology, science, finance, politics, and medicine, to name but a few), much has been researched and written in relation to risk, risk analysis, risk management, and risk-based decision-making. As this paper is about risk education, it is fitting to present a synopsis of significant research studies investigating what risk education should include and how it should be carried out. Before such a discussion however, three other topics will be explored: the communicating of risk information, prominent theories about risk analysis and decision-making, and how risk is currently being incorporated into curricular (standards) documents. Whereas an understanding of what exists within theories of risk analysis and decision-making provides a reference for what may be occurring pedagogically in risk education, a curricular analysis helps to provide a contextual perspective for how risk education is being related to curriculum content. The discussion of the communicating of risk information, which will be considered first, sheds light on what understandings students need to “survive” in our “risk-driven” world.

Communicating about Risk

Today, the assessment and management of risk, as well as the making of decisions with respect to risks involved, is found in nearly every facet of our lives, making the communication of risk-based information omnipresent. In relation to communication about risk, three aspects will be considered here: language issues, relative and absolute risk, and representational formats.

Often, in an effort to clarify risk information, complex probabilistic relationships, calculations, and values are communicated instead through the use of adjectives: “probabilities can be described fluidly with words, using language that appeals to people’s intuition and emotion” (Spiegelhalter, Pearson, & Short, 2011, p. 1394). For example, in a Rapid Risk Assessment for the Ebola virus disease (EVD) in West Africa released by the European Centre for Disease Prevention on January 30, 2015 the following communication of a risk assessment was given: “The risk of EVD being imported into the EU or the risk of transmission occurring within the EU remains low or very low due to the range of risk reduction measures that have been put in place by the Member States and the affected countries" (p. 2). Likewise, on February 11, 2015, Public Health England released a Risk Assessment of the Ebola virus disease outbreak in West Africa stating: "Despite the recent confirmation of the first Ebola virus disease (EVD) case diagnosed in the United Kingdom (UK), there is no change in the overall EVD risk assessment for the UK. While the risk of further EVD cases being imported into the UK is currently considered to be low, the risk of transmission occurring within the community in the UK is, and is expected to remain, very low due to the range of robust measures that have been put in place. There is still an expectation that a handful of cases may occur in the UK over the coming months" (p. 1). It is important to note that although both risk assessments are addressing the same two risks, EVD entering a region from elsewhere and the spread of EVD within a region, there are differences in how the assessment of risk(s) is presented, and quite possibly how those risks are determined. In the case of the Public Health of England assessment, the risk of importing EVD into the UK and the risk of transmission of EVD within the UK are considered separately, giving two definitive appearing risk ratings of low and very low, respectively. Alternatively, the European Centre for Disease Prevention's assessment considers the two forms of risk, the importing and the transmission of EVD, as a single over all risk, which provides a similar, but seemingly less definitive risk rating of low to very low. The choice of which way to evaluate and present the two official risk assessments is not justified or clarified in either document; moreover, what constitutes low or very low risk (in either case) is not elucidated (beyond the notion of a "handful of cases of EVD over the coming months" – a vague statement in and of itself), leaving interpretation of this risk assessment up to the individual. These two reports, coming from similar organizations give an example of Spieglhalter, et. al.'s (2011) claim: "the attractive ambiguity of language becomes a failing when we wish to convey precise information, because words such as 'doubtful,' 'probable,' and 'likely' are inconsistently interpreted" (p. 1394). If the use of such language is not deemed advisable in communicating in general about risk, consideration needs to be given to how to otherwise communicate the information, and how to educate people to interpret such communications.

Also related to language is the concern raised by some researchers (e.g., Martignon & Krauss, 2009; Pratt, Ainley, Kent, Levinson, Yogui, & Kapadia, 2011; Till, 2014) over the deceptive (intentional or otherwise) use of relative (perceived) risk, rather than absolute risk, within communications. Relative risk, which is a comparative statement of the change in risk or a comparison of two different risks, can lead to grievous misunderstandings. Consider, for example, a dog food commercial that advertises that their new product has 50% more protein than that of their old product. For many pet owners, a higher protein diet is desirable and the relative risk (positive, in this case) of 50% can easily be interpreted as a significant increase, but in fact, the absolute risk, or how significant the increase really is, depends upon how much protein was in the original product.

In considering the communication of risk information, a final focus of research is of importance in this discussion: how individuals understand (or misunderstand) different representations of information related to risk. Till (2014) notes: "Findings of cognitive psychologists reveal evidence that the format of representation is crucial for understanding the real harm or chance of different options in situations of uncertainty … Frequency formats are much better processed by the human mind than ratios" (p. 84). Till also argues that "Since risk-related data may be emotionally loaded, it is convenient to use representation formats that are objective, unbiased and easy to grasp for a wider public" (84); however, he also warns that "mathematical formats like ratios, fractions, percentages or decimals" can be misunderstood despite the visual or graphical representation used. Likewise, Spiegelhalter, et. al. (2011) discuss similar issues arising from the formatting and representational model of risk-related data and information. Clearly, the issue of how to represent risk information is a complex one, with further research in how to clarify such representations both in terms of format and public awareness and understanding being needed.

Theories Related to Risk: Analysis and Decision-Making

Much discussion about risk analysis and decision-making starts with the emergence of the term 'economic man' in the early 20th century. In the late 19th century, 'economic man' was a derogatory term used by those in opposition to John Stewart Mill's assertion that in political economy, man seeks to gain as much as possible with the least effort or loss (Persky, 1995). Moving into the 1900s, 'economic man' became re-envisioned as describing man as a being aware of all possibilities and choices in a situation and therefore capable of making decisions that maximize his advantages, minimize his disadvantages, or both. The theory of the 'economic man' soon led to the development of the 'subjective expected utility' (SEU) model. This model proposes that for any possible plan of action, there exists a set of hazards to which numerical values representing the impact of each particular hazard can be assigned. The product of the probability of a hazard occurring and the assigned value of impact for the hazard is defined as the risk of the hazard. The risk of a plan is the sum of the risks of the hazards within the plan, and the plan with the lowest risk is the best decision to be made. By the 1940s, doubts about the 'economic man' and of the SEU model were well known: the complexities involved in any one risk-based decision-making task would prevent either theory being fully realized (Kent, Pratt, Levinson, Yogui, & Kapadia, 2010). In response, researchers (such as Kahneman, Simon, Slovic, and Tversky) proposed instead that man works within a 'bounded rationality', a reality in which not all is known, and even what is known need not always be considered. Simon (1959), shifted the focus from maximizing (minimizing) to satisfying (i.e., 'economic' man to 'satisficing' man), wherein acceptable or adequate choices are made based upon limited knowledge of the full reality of a situation. From that point, researchers (such as Kahneman, Slovic, and Tversky) undertook defining and investigating different heuristics and biases "which reduce the complex tasks of assessing probabilities and predicting values to simpler judgemental operations In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (Tversky, & Kahneman, 1974, p. 1124). The heuristics and biases proposed were seen as the mechanisms used by the 'satisficing man'. As will become evident shortly, in many ways the reasoning behind the 'economic man' and the 'satisficing man' and the associated theories, can also be seen to play out within recent theories and perspectives on risk education, including where risk should be housed within curricula and how it should be approached pedagogically.

Other Data and Modelling Concerns

With all the different types and styles of risk assessments being made, guidelines are often determined to define "acceptable" risk, but as with the Ebola virus example above, what is to be considered within one establishment as acceptable risk is not likely to be consistent with another, even in the same field. One need only to consider the current issues of water use, transportation of gas and oil, or fracking to realize that it is very possible that one person's acceptable risk is another's unacceptable risk. Throw "big data" into the mix, and the determination of risk becomes more complex. As Spiegelhalter (2104) states, "Big data means that we can get more precise answers… But this apparent precision will delude us if issues such as selection bias, regression to the mean, multiple testing, and overinterpretation of associations as causation are not properly taken into account… Serious statistical skill is required to avoid being misled" (p. 265). Tim Palmer, a physicist specializing in weather prediction and climate change, adds more uncertainness to the power of big data in risk assessment: "The truth is that the level of detail in the models isn't really determined by scientific constraints… It is determined entirely by the size of the computers" (Macilwain, 2014, p. 1222). Unlike Palmer's concern about our ability to use big data meaningfully within technological constraints, Smith, an economist, referring to climate change models based upon past mean-temperature changes over 1-10 years, argues: "The question is, when will we have significantly better quality information than we have today? I think we may have our answer from the climate before we get it from the physics'" (Macilwain, 2014, p. 1223). Thus, questions about risk related to big data need to take into account more than the issues that Spiegelhalter (2014) noted: the capacity of computers and the ability for meaningful and relevant data to be collected for analysis must also be scrutinized. Meanwhile, risk assessments and management strategies are in constant demand for those seeking to make well-informed decisions.

Current Curricular Inclusion of Risk and Risk-related Concepts

Since risk impacts our lives in so many ways, it is not surprising that many educational systems and researchers are looking for ways to embed risk assessment and management and risk-based decision-making into the K-12 school system. Speaking about the UK, Pratt, Levinson, Kent, & Yogui (2011) comment: "In modern society, risk permeates decision-making at both personal and policy levels, a fact now being recognized in curricula" (p.1), including "Personal, Social, Health and Economic (PSHE) Education, Citizenship, Science and to a lesser extent in Mathematics" (Pratt & Yogui, 2010, p. 1). The word "risk" appears in at least one of the key stages for the above-mentioned curricula, except, that is, Mathematics (hence "to a lesser extent"), where 'risk' is not explicitly stated, but is implicitly connected to some of the content outlined.

PSHE is different from the other courses mentioned because it is a non-statutory course. Instead of a curriculum (or standards), PSHE has a guidance document that outlines the intent of the various key stages, which is meant to guide the development of school-based curricula for the courses. Within this guidance document, the following statement is made: "we expect schools to use their PSHE education program to equip pupils with a sound understanding of risk and with the knowledge and skills necessary to make safe and informed decisions" (Department of Education, Sept. 11, 2013). Consequently, as a non-statutory course, the defining of risk and what knowledge and skills are necessary to make safe and informed decisions is left in the hands, for the most part, of individual schools.

The remaining courses sited by Pratt and Yogui (2010) are statutory courses and thus have specific curricula (standards) for the different key stages or key stage pairings. For example, in Citizenship, at key stage 3, pupils are to be taught about "the functions and uses of money, the importance and practice of budgeting, and managing risk" (Department of Education, September, 2013, p. 2). This is a much narrower focus on the type of risk (specifically, financial) than given in PSHE, but as noted about the PSHE course, how such risks are to be determined or managed is not discussed. Similarly, in key stage 3 Science, pupils are to be taught attitudes towards working scientifically, including to "evaluate risks" (Department of Education, 2013, p. 4), yet there is no indication of what risk and its evaluation entails. The same is true in key stage 4 Science, where it states that as part of their development of scientific thinking, "students should be taught so that they develop understanding and first hand experience of evaluating risks both in practical science and the wider societal context, including perception of risk" (Department of Education, 2014, p. 5). It is likely that such notions are addressed in approved resources, raising the question, what should the resources include about these topics?

As noted previously, risk, by name, is not mentioned in any of the key stages of the Mathematics curriculum within the National Curriculum in England. There is, of course, much content related to probability and statistics, which inherently would seem to connect itself to risk assessment and management as well as risk-based decision-making. This content, taught (as described in the document) through the pupils learning about working mathematically, such as selecting "appropriate concepts, methods and techniques to apply to unfamiliar and 'non-routine problems' interpret their solution in the context of the given problem" (Department of Education, 2014, p. 6), could quite reasonably be interpreted as including risk-related learnings.

A search of the Common Core State Standards Initiative (2015) finds two standards in which risk-related analyses and decision-making are the focus: CCS.Math.Content.HSS.MD.B.5 -- "Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values" and CCS.Math.Content.HSS.MD.B.7 --"Analyze decisions and strategies using probability concepts (e.g., product testing, medical testing, pulling a hockey goalie at the end of a game)". In both instances, the risk assessment being considered is grounded in probability and probabilistic reasoning. As only Mathematics and English Language Arts have Common Core State Standards, the concept of risk within other subject areas will vary according to the curricula or standards adopted in a particular state or region.

As just one example of what is happening in the United States within the other subject areas, consider the Content Standards for California Public Schools (California State Board of Education, 2014), which mentions risk in all but 7 of the documents posted on the site (specifically, English Language Arts, Building and Construction; Energy, Environment, and Utilities; Engineering and Architecture; Fashion and Interior Design; Manufacturing and Product Development; and World Languages). In many cases, "risk" in the content standards is used in a discussion of at-risk students; however, specific references to risk assessment and management, as well as risk-based decision-making, are given. For example, the *Health Education Content Standards for California Public Schools, Kindergarten Through Grade Twelve* (California Department of Education, 2009) document includes, in the rationale for Standard 4: Interpersonal Communication, the statement: "The ability to appropriately convey and receive information, beliefs, and emotions is a skill that enables students to manage risk, conflict, and differences and to promote health" (California Department of Education, 2009, p. viii) and Standard 7: Practicing Health-Enhancing Behaviors states: "All students will demonstrate the ability to practice behaviors that reduce risk and promote health" (p. viii). Through all grades, the Health standards frequently reference students learning about ways to reduce particular risks and to avoid risky behaviors. Alternatively stated, the focus is mainly on identifying risk factors and options related to risky situations or the awareness of and management of risk. In addition, decision-making related to risk is also considered (mainly from the perspective of reducing risk). Analysis of risk does not appear to be a primary focus; however, the focus on knowing about risks would be a primary step towards analysis of them. Within the *Grades 9 – 12 by Disciplinary Core Ideas* (Schweingruber, Keller, & Quinn, 2012) document for Grades 9-12 Science, learning related to ideas of risk considers the "limitations on the precision of data (e.g., number of trials, cost, risk, time), and refine the design accordingly" (p. 47). This statement harkens back to some of the concerns expressed by Spiegelhalter (2014), and Palmer and Smith (Macilwain, 2014) and it is likely that such a consideration would include an analysis of the risk involved in the identification and collection of data and decisions being made upon that analysis.

Like the United States, Canada also does not have a single unified set of curricula; however, the four Western provinces, the three territories, and some of the Atlantic provinces are all using the Western and Northern Canadian Protocol (WNCP) Common Curriculum Frameworks for K-9 and 10-12 Mathematics (CCF) as the foundation for their respective provincial and territorial mathematics curricula. For some of these jurisdictions, the curricula documents used are identical to the WNCP CCFs (save the cover page), while in other jurisdictions, the same mathematical content is presented in different ways in order to better communicate the specific initiatives and priorities of that jurisdiction. In some (rare) cases, content is added or deleted from the WNCP CCFs at the jurisdictional level; however, the overall agreement between the different curriculum documents is still very high.

Within the WNCP documents, risk is specifically mentioned in upfront matter, which gives a brief synopsis of the pedagogical and mathematical beliefs and processes that the document is grounded upon. These references to risk all relate to the importance of encouraging and supporting intellectual risk taking within mathematics: "Creating an environment where students openly look for, and engage in, finding a variety of strategies for solving problems empowers students to explore alternative and developed confident, cognitive mathematical risk-takers" (WNCP, 2008, p. 8). However, this is not to say that risk analysis and risk-based decision-making does not get consideration within the curricular documents. Consider, for example, the first outcome in Grade 12 Workplace and Apprenticeship Mathematics (Saskatchewan Ministry of Education, 2013b): "Analyze and interpret problems that involve probability", and specifically, indicator 1.6 for this outcome: "Explain, using examples, how decisions may be based on a combination of theoretical probability calculations, experimental results and subjective judgments" (p. 26). It can be easily argued that such decisions may well, in fact likely would, be in relation to situations involving risk. A similar outcome, in the course Grade 12 Foundations of Mathematics (Saskatchewan Ministry of Education, 2013a), " Interpret and assess the validity of odds and probability statements" has the indicator "1.5 Explain, using examples, how decisions may be based on probability or odds and on subjective judgement" (p. 71), which is also easily connected to risk assessment and management, as well as risk-based decision-making. Thus, within these two courses, students should be engaging in risk analysis and risk-based decision-making through the consideration of both objective and subjective knowledge. It can be argued that these two indicators give a bit more direction towards what students might learn when reasoning about risk than what was found in previously discussed curricular documents, namely probability, odds, and subjective judgments in decision-making; however, where it is obvious what objective knowledge the students are to consider, i.e., probabilities and odds, what kinds of subjective knowledge to be considered (or if it is open-ended) remains undefined.

Risk Education

With the above insights into some of the more common ways that risk, risk analysis, risk management, and risk-based decision-making have been incorporated into a selection of (Western) curricular and standards documents, thoughts now turn to what is being said and researched about in relation to risk pedagogy. Until recently, when risk started to impact people's daily lives in a myriad of ways, risk education, particularly at the K-12 levels, was generally not being considered, and thus, was not a focus of research either. In the field of risk education, links between existing theories and research on risk and decision-making in other domains are often forged along with considerations of new ideas and relationships. Within this paper, two particular aspects will be emphasized: the assumptions, beliefs, and theories supporting the approaches taken, and the approaches themselves. Given primary focus herein is the work of two sets of researchers: Krauss and Martignon; and Ainley, Kapdia, Kent, Levison, Pratt, and Yogui.

Research by Krauss and Martignon. The research in risk education by Martignon and Krauss (2009) focused on developing within students a "chain of competencies that make up good decision making for informed consent in basic domains of modern life like those of medical and investment decisions" (p. 229). Further, recognizing that "while most of the mathematical training of secondary school tends to be soon forgotten by those who do not pursue a career requiring further mathematical tools, mathematical competencies acquired until the ninth or tenth year of age appear to remain robust and unaltered during subsequent life" (p. 228), grade four students were selected to be participants in the study. In carrying out their study of the grade four students' development of the particular competencies, Martignon and Krauss turned to research findings demonstrating that hands-on activities and tools can strengthen learning and sought ways to incorporate such strategies into their pedagogical design.

Martignon and Krauss (2009) also provide an explanation of the major assumptions behind their intent of the study, and in particular, underlying the chain of competencies ultimately investigated:

> Stochastic literacy is a necessary condition for enlightened decision making in an information-based society. Becoming conscious that judgments about our fellow human beings and about nature should often be based on probabilistic rather than strict logical implications can reduce the impact of prejudices and stereotypes. Moreover, an understanding of probabilities can shape our decisions allowing us to assess possible risks associated with our actions. In fact, good modelling of risky situations can sustain our cognitive and emotional perspective on personal and collective affairs, reducing our anxieties and guiding our informed consent (p. 227).

A number of key ideas emerge from this statement, the first of which is the claim that enlightened decision-making is dependent upon a significant understanding of stochastics. Second, the claim is made that probability-based reasoning can reduce the impact of prejudices and stereotypes. A third assertion is that probabilistic knowledge can support risk assessment, and finally, that probabilistic modelling, done well, can mitigate the impact of otherwise held cognitive and emotional perspectives by moderating anxiety and providing support for strong decisions. In other words, probabilistic and statistical knowledge and reasoning are being centered as a (if not the) key player in risk education; moreover, such learnings can reduce (perhaps even eliminate) negative results of more affective responses, such as anxiety, prejudices and stereotypes. As such, the focus of the study is on the development of rational competencies and indirectly, the diminishing of the impact of affective reasoning.

The series of competencies that this study investigated are intended to act as a "tool box for decision heuristics in the bounded rationality paradigm" (Martignon, & Krauss, 2009, p. 229) for students in the classroom and in their future lives. Within this bounded rationality paradigm, Martignon and Krauss argue that, like the satisficing man, people "combine elements of basic Bayesian reasoning with effective decision heuristics" (p. 229) that they have selected to be of importance. The purpose of the play-based activities in the study are thus to give the students the tools necessary to make rational appraisals and decisions that are both effective and efficient (within a bounded rationality paradigm).

In the study, through the use of cards and tinker cubes, the teaching of the grade four students begins with the investigation of making logical inferences within the context of if-then statements. However, "Preparing children for decision-making practices requires training in making inferences, not just strictly logical but also, most importantly, probabilistic ones" (Martignon, & Krauss, 2009, p. 231), and so the students move next onto activities involving "conditional implications" (p. 232) as a consequence of the inclusion of conditional probabilities. From there, the students engage in activities that involve the comparing of proportions. The researchers explain: "[students] need to understand conditional probabilities for determining the validities of features and they need to make comparisons between different validities of features for establishing rankings among features… These competencies are at the core of risk assessment" (p. 231). Further, Martignon and Krauss explain: "The comparison of proportions is essential in comparing feature validities and for assessing risks" (p. 232). The research, and the experimental teaching, for this study ends with the students playing a game in which comparing of risks can be used to determine which of a number of possible moves is the least risky. Overall, the ordering of the activities is described as "a 'historic trajectory'… from logic to probability" (p. 238), giving the students a historically accelerated experiential learning of decision-making and reasoning with risk.

Research by Ainley, Kapdia, Kent, Levinson, Pratt, and Yogui. Foundational to their study of risk education, Kent, Pratt, Levinson, Yogui, and Kapadia, (2010) explain that: "Going beyond the idea of risk in statistical theory, we are trying to understand how personal values and models influence thinking about risk and the process of decision-making, and the implications of this for classroom practice" (p. 1). Thus, different from Martignon and Krauss (2009), this group of researchers considers the role of affective (emotional) knowledge and responses in decision-making and risk by making "An initial assumption ... that decision-making involves the coordination of different kinds of information, based on quantitative models and personal value systems and judgements" (p. 1). This study also differs from the one previously described in that it is focused on a different category of research participants: mathematics and science teachers with an extension to one group of high school students. The selection of mathematics and science teachers as participants is related to the previous discussion of risk within curricular (standards) documents, as well as a recognition that often the socio-scientific aspects of risk and decision making are seen as part of science education, while the stochastic aspects are seen as part of mathematics education. The participants were paired (one mathematics teacher with one science teacher) in the study to shed light on this assumption, but also to see the result of having both approaches working together as an attempt to inform where and how risk might best be brought into curricula in general. Overall, it was hoped that the participants would come to a better understanding of risk and decision-making processes and features through their involvement in the study; an understanding which might then be taken back into the respective teachers classrooms. A final part of the purpose of their study was to investigate how the results of the study aligned with existing research and theories about risk and decision-making. In particular, a focus was given to investigating the prevalence and strength of the application of the priority heuristic by the participant teachers (see Pratt, Levinson, et. al., 2011); however, in this paper the focus will remain on the reasoning behind the design of the decision-making software tool that was developed and implemented and on the choice of participants.

Reflecting upon the various theories and research of risk, Kent et. al. (2010) argue: "we think it is clear that students cannot be educated to think about risk only from a heuristic basis. There is a need for a systematic quantified analysis of some kind" (p. 2), which the team decides to attempt through the creation of a set of software tools entitled Deborah's Dilemma. The first characteristic that should be identified about Deborah's Dilemma is that the researchers attempted to create the tools to deal with a particular real-world (like) situation. This decision is supported by the work of Gal (2012) who argues that, with respect to risk education, "The content of instructional tasks should not be limited to traditional examples taken from games of chance or based on artificial problems, but the kinds of everyday contexts where probability plays a role and where adults encounter uncertainty and risk" (p. 6). Deborah's dilemma is she has been told by a doctor that she needs to have back surgery. The user of the software tools is provided contextual information regarding minor and major hazards related to the surgery, and aspects of Deborah's life that are impacted by the health of her back. Based on this information, the participants are asked to recommend to Deborah whether or not she should go ahead with the surgery. As is often the case in real life, the information provided is at times inconsistent and incomplete, as well as from a variety of sources.

The software package developed has three tools for the user to engage with: the 'Operation Outcomes' tool, the 'Painometer' tool, and the 'Risk Mapping' tool. The Operation Outcomes tool allows the user to input various probabilities obtained from the information provided regarding the possible outcomes of having the surgery (successful to death, inclusively). The tool uses the probabilities entered by the user to run simulations of the surgery occurring. The user then requests the tool to carry out single or multiple runs (surgeries) and an outcome (or set of outcomes) is generated by the program. Because the participants chose which hazards to consider in the use of this tool, their affective responses were necessarily part of their decision-making.

The Painometer tool allows the teachers to "control Deborah's level of pain tolerance, the amount of work and domestic/leisure/sport activity that Deborah does and the pain intensiveness of each, assuming that some types of work and sport would worsen the pain and others would relieve" (Pratt, Ainley, Kent, Levinson, Yogui, & Kapadia, 2011, p. 330). In addition, the participants could add activities to the painometer to esplore "whether a balance could be achieved between pain-inducing and pain-relieving activities, so that Deborah might manage her pain within tolerable limits from day to day" (Pratt, Ainley, et. al., 2011, pp. 330-331). Because the participants controlled the level of pain assigned to each activity (compared to a standardized tolerance level), the Painometer tool allowed additional aspects of their affective responses to be accounted for.

An important confirmation for the researchers from an initial trial of the study, where only the Operation Outcomes and Painometer tools were used by the participants, was that the participants were struggling with bringing the two aspects influencing their decisions (the possible outcomes of the operation and the levels of pain associated with daily tasks without the operation) because their analyses were taking place in two disconnected tools. As a result, the researchers implemented the 'Risk Mapping' tool – a graphical modeling tool of decision boxes and hazard boxes. Colour coding of the boxes allowed the participants to represent the level of risk associated with the various decisions and hazards (as opposed to numerical probabilities). The researchers explain: "Whilst the mapping tool does enforce the association of impact and likelihood with each hazard, we did not enforce any model for how these relate to 'level of risk'. It was exactly at this point where we hoped users would express their personal models for the situation, providing us with a window on their thinking about risk" (Kent, et. al., 2010, p. 4), once again assuring the possibility for both rational and affective reasoning to emerge. With this third tool, the participants were found to be able to better coordinate their thinking about the hazards and their thinking about the impacts.

With this brief overview of pertinent literature and research related to risk, risk analysis, risk management and risk-based decision-making, the stage is now set for the introduction of the theoretical framework for this paper. As mentioned previously, this theoretical framework is based upon two distinct worldviews, an Indigenous worldview and the Traditional Western worldview, and analysis occurs within an ethical space. The introduction to the theoretical framework begins with a discussion of broad understandings and considerations necessary when working within a worldview framework.

Theoretical Framework

To carry out a theoretical analysis of risk education from the perspective of the two aforementioned worldviews, it is important to acknowledge the definition of worldview being used in this paper and to clarify with whom each of the worldviews are associated. Worldview is a term that has been defined in many ways, depending upon the context and resulting purposes in which it is to be used. For example, when considering worldviews within the context of different religions, one might define alternate religious worldviews according to answers and approaches to the big questions of life (e.g., what is the purpose of life, where did life come from) and how and where the answers to those questions can be sought or found. Regardless of the context (religious, economic, political…) however, all worldview descriptions ultimately consider (and answer) two questions: "what kinds of knowledge are of value" and "what ways of knowing are valued". It is this less context dependent and more holistic way of defining individual worldviews that is used in this article.

Equally important to understand about the worldviews and their descriptions is that the particular names of the worldviews should not be thought to mandate inclusion or exclusion of any individual from said worldviews on the basis that he/she is or is not considered a(n) 'Western' or 'Indigenous' person. The names of these two worldviews comes from observed commonalities within the named group in general, but it is not to be concluded that an individual holds said worldview solely on the basis of their membership in the larger group (e.g., a person who identifies as being of 'Western' heritage need not be grounded within the Traditional Western worldview). Conversely, holding a particular worldview is not restricted by what group one is a member of (e.g., a person identifying as 'Indigenous' may hold the Traditional Western, or any other worldview).

Some concern may also arise over the use of different articles ('the' and 'an') proceeding the two worldview names ('the' Traditional Western worldview and 'an' Indigenous worldview); however, this is done with purpose. As will become evident shortly, the Traditional Western worldview is one in which little (or no) variation is possible because of the emphasis on singularity of truth, knowledge, and ways of knowing. As a result, it is distinguishable from other possible Western worldviews. Hence, 'the' Traditional Western worldview is 'a' Western worldview – one of many possibilities. 'An' Indigenous worldview, as again will become evident shortly, does not preclude variability within the worldview. An Indigenous worldview, as presented, is not representative of any one particular Indigenous group's worldview. Rather, it is an overarching worldview comprised of commonalities amongst the ways of knowing and kinds of knowledge valued by different Indigenous groups and peoples around the globe. This worldview has within it the flexibility to fit the variance of specific worldviews found amongst Indigenous groups and peoples, thus it is really one of a range of Indigenous worldviews, hence 'an'. With these broad understandings and considerations in hand, brief introductions to the Traditional Western worldview and an Indigenous worldview, as understood for the theoretical framework for this paper, will now be provided.

The Traditional Western worldview

The Traditional Western worldview has its own "ontological, epistemological, sociological, and ideological ways of thinking and being" (Kovach, 2009, p. 21) that can be distinguished from those of other worldviews, such as an Indigenous worldview. The Traditional Western worldview has as its foundation at least five defining characteristics, which, in a rational way, confirm and strengthen each other (Absolon and Willett, 2005; Kovach, 2009; Ermine, 2007; Little Bear, 2000; Meyer, 2003; Schelbert, 2003). Perhaps most important to the Traditional Western worldview is the belief that knowledge of value is "linear and singular, static, and objective" (Little Bear, 2000, p. 82) in nature, resulting in the valuing of one correct answer to any problem or question; further, this answer cannot be proven false or replaced by alternative answers. In effect, this characteristic of the Traditional Western worldview places an emphasis upon the defining of dichotomous relationships. In addition, the Traditional Western worldview holds that there is one correct way to achieve these answers (at least within each category of context considered). The valuing of the linearity of knowledge naturally gives rise to the defining and perceiving of hierarchies of knowledge, giving rise to specialization and superiority and authority of both knowledge and knower. Seeking of more knowledge, then, is a way to become more specialized, more superior, and more authoritative. Within the traditional worldview, new knowledge is sought for the sake of the knowledge itself. These hierarchies, based upon linearity, help to eliminate the possibility of alternative answers or solution strategies – as one moves up a knowledge ladder, the newest knowledge level is given authority over all previous levels. The Traditional Western worldview holds rational thought and the scientific method as crucial in the obtaining knew knowledge, or stated more explicitly, knowledge of value comes from observation that is done in total isolation of all other factors and is based upon measurable data. The result is that the notion of truth being linked to measurability, and knowledge of value being related to physical objects and processes that are external to the individual and that are in isolation of other intruding influences. Directly related to the scientific method and rational thinking are the notions of the compartmentalization and categorization of knowledge into small components which are also important within the Traditional Western worldview. Through compartmentalization and categorization, knowledge is believed to not only be emerging in an isolated way (not interfered with by extraneous factors), but it is also easier to accurately measure. Together, these processes confirm not only validity, but absolute truth; further, that absolute truth is known by experts who are attributed with absolute authority. Another consequence of the rationality, measurability, compartmentalization, and categorization of knowledge of value is that it is also abstract knowledge – void of context and dissected into small parcels of truth. Finally, as a consequence of the truth of the valued knowledge never changing, the Traditional Western worldview values the preservation of this knowledge in written, an abstract (symbolic) and permanent form in which the truth is captured for perpetuity.

An Indigenous Worldview

In an Indigenous worldview there are number of defining characteristics, all explicitly and implicitly linked to one main characteristic: relationships (Absolon 2010, Barnhart & Kawagly, 2005; Canadian Council on Learning, 2007; Ermine, 1995; Hogan, 2000; Kovach, 2009; Leavitt, 1995; Little Bear, 2000; Meyer, 2003; Youngblood Henderson, 2000). The main encompassing characteristic of this worldview, the establishment and maintenance of relationships with all of creation (including people, the earth, the spirit world, and the cosmos), is foundational to all knowledge of value and ways of knowing. It is through the establishment and maintaining of relationships that knowledge worth knowing emerges. As a result, knowledge of value in an Indigenous worldview also contributes towards the perpetuation and strengthening of relationships. Moreover, the relationships that are gained and maintained in the seeking of knowledge come from all ways in which a relationship can exist: physical, emotional, spiritual, and intellectual. In an Indigenous worldview it is important that knowledge is intrinsically and extrinsically connected to the place from which it came – the knowledge is in relationship to place. Knowledge that only exists in abstraction where it is decontextualized and disassembled into its constituent parts no longer carries with it the connection to the relationship(s) for which and through which it was created and, as a consequence, such knowledge becomes less valuable. In addition, since knowledge is created and shared through relationships that are not just intellectual in nature (but also emotional, physical, and spiritual), the restriction that knowledge of value comes from objective sources and rational processes is not found within an Indigenous worldview, rather subjective knowledge is often viewed as just as valuable, or even more valuable, than objective knowledge. Personal experience and intuition (from the past, present, and future) are also considered valid sources of knowledge. Moreover, within an Indigenous worldview, valuable knowledge is understood to be different for different people, and the diverse ways in which these different knowledges can be created are also of value. This appreciation for diversity in knowledge and ways of knowing also derives from the importance of relationships, where getting to know and appreciate the uniqueness of who or what one is in a relationship with is crucial for the relationship to be authentic, meaningful, and productive. It is as a result of this characteristic of an Indigenous worldview that the knowledge that is created and valued is able to remove the sense of dichotomy between what might otherwise seem to be diametrically opposed ideas, such as good and evil. Instead of being polar opposites, the two states become two parts of the totality of being. This means that within the Indigenous worldview, alternate answers and alternate solutions (and strategies) to a problem are not only recognized, but valued for the wisdom that they bring to relationships. As a significant part of any relationship is seen to be the act of giving back, sharing is also an important part of an Indigenous worldview, as is the use of the knowledge gained through relationships to strengthen those relationships and create new ones. Knowledge is not sought for the sake of knowledge, but for the sake of how it can contribute to the wellbeing of the whole (including self, family, community, the earth, the universe and the cosmos). Since knowledge varies as relationships change and emerge, knowledge of value is most often kept and shared through oral traditions, allowing the sharer of the knowledge the flexibility to adjust the knowledge for the relationships into which it is being brought; sharing of knowledge and giving back through knowledge is therefore sensitive to relationships, keeping relationship central to the valuing of knowledge. This oral tradition and the knowledge contained therein contribute to the traditional (cultural) knowledge of the Indigenous people and the place and story in which the knowledge originated is considered part (the story) of the knowledge. Without place, and thus story, the knowledge has no value.

The Meeting of the Two Worldviews: Ethical Spaces

Much tension can (and historically, has) easily formed between the two worldviews described. For this reason, the analysis that follows will be considered from within an ethical space, that is, the space that exists between two different knowledge systems. Ermine (2007) explains that such a space "is initially conceptualized by the unwavering construction of difference and diversity between human communities" (p. 194), just as the descriptions of the Traditional Western worldview and an Indigenous worldview has. By considering these differences – which define the uniqueness of each worldview resulting from "distinct history, knowledge tradition, philosophy, and social and political reality" (p. 194) – and then contrasting the worldviews in a non-judgmental way that creates and allows for engagement in the ethical space between the two cultures. Therefore, analysis using the theoretical framework of the two worldviews carried out within an ethical space is not meant to evaluate or rank the two worldviews; rather, it is an opportunity to better understand what is currently under consideration in research pertaining to risk education and to hypothesize where else the research might head.

In order to engage in such an ethical space, Ermine (2007) holds two criteria must be met: the context under consideration must be unethical, and there needs to be cross-cultural concerns. With respect to ethical spaces, Ermine defines ethics as "the capacity to know what harms or enhances the well-being of sentient creatures" (p. 195), so an unethical issue would be one that has caused, or has the real potential to cause harm, to which it is contended in this paper, that risk education is focused on the potential for harm, and that risk education has the potential to influence (positively, negatively, or not at all), students' ability to negotiate situations and events involving risk. Therefore, risk education can be considered an unethical issue.

The second criteria of the meaningful entrance into an ethical space is that there needs to be cross-cultural concerns related to the issue of interest. Of course, risk is a topic or issue, which knows no cultural bounds in the broadest connotations; however, feminist theory, critical race theory, theories of decolonization, and other post-modern and post-structural theories acknowledge a multitude of cross-cultural concerns that relate to risks and disparities in risk with respect to culture. For example, currently in Canada, the risk-based cross-cultural concern of missing and murdered Indigenous women (see Kappo, 2014 for a brief overview of the cultural nature of this issue) is a prominent issue of interest for many.

With an understanding of the two worldviews and of an ethical space in which they can be used to analyze risk education, the stage is set to consider what aspects of the research of risk education align with aspects of either of the worldviews. Further, consideration can also be given to those aspects of either worldview which, so far, have not entered into the investigation of and theorizing about risk education. These ideas can then be removed from the ethical space to see if they might be meaningful and plausible within risk education itself.

Analysis

The theoretical framework of the two worldviews (the Traditional Western and an Indigenous) is now brought to bare, within an ethical space, upon the aspects of risk and risk education discussed in the literature synopsis above. It cannot be overstated that the purpose of this analysis is not to judge what has been done within existing research, but to present a new way of thinking about that research in light of the kinds of knowledge and ways of knowing that are being promoted or highlighted, and thus valued. Likewise, it provides an opportunity to identify other possible kinds of knowledge and ways of knowing that have been either overlooked, ignored, or denied. Neither of these analyses is a judgment on choices made or not made, but rather a framing of what is and what else might be possible or considered. The analysis begins by considering what ways of knowing and kinds of knowledge are already being valued, both implicitly and explicitly, within the existing research, then turns to what is not present.

Valued Knowledge and Ways of Knowing In Communications and Risk Research and Theories

Both the topics of communicating about risk, and theories of risk analysis and decision-making are inconclusive regarding how or if they align with aspects of either of the two worldviews in relation to the kinds of knowledge and ways of knowing being valued. In many communications about risk, the kind of knowledge and ways of knowing that are considered of value are not typically presented, only the conclusions made as a result of the particular positioning. For example, the two EVD risk assessment communications do not indicate what information the risk assessments are based upon, nor is the way in which that knowledge is known described. Assumptions, based upon past experiences and possibly even gut feelings, might cause one to lean towards a particular conclusion (e.g., this is rational knowledge based on scientific methods), but in reality only in-depth research of each particular circumstance could say for sure. The use of adjectives in risk communications is equally disagreeable towards the undertaken analysis. Although one might be positive that the risk that is being communicated about is based upon scientifically backed rational thinking, along with the hierarchical and compartmentalizing nature (e.g., very low, low, moderate, and so on) of their use (thereby aligning with the Traditional Western worldview); the subjective possibilities for interpretation of the terms would seem to lean more towards the kinds of knowledge and ways of knowing that are valued within an Indigenous worldview. Likewise, the mathematically theoretical nature of the SEU model would seem to indicate that linear, static, authoritative, and objective knowledge and ways of knowing are being emphasized, yet it also inherently must involve subjective reasoning related to what features of the risk are or are not considered (or identified). It is within the more applied aspects of the literature synopsis (i.e., risk in curricular/standards documents and risk education) that the theoretical analysis proposed above takes on more relevance and less ambiguity.

Valued Knowledge and Ways of Knowing In Curricular (Standards) Documents

Like above, in some of the examples provided regarding inclusion of risk and decision-making within curricular (standards) documents, the ways of knowing and the kind of knowledge being valued are not always evident; however, in some there are more easily identifiable indications. For example, the UK's Citizenship document, managing risk is tied to very factual and procedural ideas: functions and uses of money and budgeting. It would seem then, that likewise it would be expected that the way students should know about managing risk would also be in a routine-oriented, hierarchical and rational-based way. Consequently, the emphasis would also appear to be on knowing of facts and procedures. Thus, both the ways of knowing and the kinds of knowledge valued by this inclusion of risk within Citizenship could be said to align closely with the kinds of knowledge and ways of knowing valued within the Traditional Western worldview. Similarly, within California's standards document for health education, there are also (likely) indications of valuing rational and scientific knowledge and ways of knowing, such as in the practicing of risk-reducing behaviours. In so saying, the assumption is that the preferred risk-reducing behaviours would be presented to the students as rational and factual knowledge by an authority figure (the teacher), all of which is in direct alignment with the valuing of the Traditional Western worldview. However, the rational for Standard 4 document what is valued shifts from strictly static and objective knowledge to knowledge that includes beliefs and emotions which are of value within an Indigenous worldview. Moreover, by including beliefs and emotions, not only is what kinds of knowledge valued expanded – so too are the possible ways of knowing expanded to include emotional ways of knowing and possibly others (such as spiritual, physical, intuitional, or experiential). The two examples from mathematics curricula from Saskatchewan add another twist to the analysis since despite the absence of the word risk within the indicators or outcomes, it is reasonable to assume that risk would be a consideration within the decision-making, if only on the basis that the decision-making is within the context of probability education. In these two examples, both objective and subjective forms of knowledge are to be considered. In the Workplace and Apprenticeship Mathematics course it is made clear that the objective knowledge might originate from theoretical calculations or from experiments, while in the Foundations of Mathematics course no such clarification is made. With respect to the kinds of objective knowledge and subjective knowledge that are to be considered, neither course offers further clarification. Thus, it can be contended that these two examples of outcome indicators are specifying that both rational and scientific knowledge and ways of knowing as well as subjective knowledge and ways of knowing are to be considered within decision-making, and thus valued. Consequently, the ways of knowing and kinds of knowledge valued aligns with the Traditional Western worldview and at least some of an Indigenous worldview.

Valued Knowledge and Ways of Knowing in Risk Education – Martignon and Krauss

The discussion of risk education, because of greater breadth and depth of information available, will allow for the most comprehensive analysis of the valuing of kinds of knowledge and ways of knowing. For example, consider the stated assumptions for Martignon and Krauss' (2009) study. First, "Stochastic literacy is a necessary condition for enlightened decision making" (p. 227) is a statement demonstrating the valuing rational knowledge and ways of knowing. Likewise, the implication of stating that probabilistic reasoning can impact prejudices and stereotypes is also conveying a strong message about the importance of rational knowledge and ways of knowing over prejudices and stereotypes, which are often argued to be the products of emotional rather than rational knowledge. Of course, the reverse argument can be made that many prejudices and stereotypes are proposed and defended through what, at least on the surface, appears to be rational and scientific evidence. Finally, as informed consent is being couched within good modeling (undoubtedly based upon probabilities) the implication would be that such informed consent also comes from rational and scientifically produced knowledge. Thus, the kinds of knowledge and ways of knowing supported by the assumptions of Martignon and Krauss' (2009) would best align with those of the Traditional Western worldview.

Considering the design of Martignon and Krauss' (2009) study there are a number of ways in which the valued knowledge and ways of knowing can be directly related to those of the Traditional Western worldview, starting with the ordering of the activities. First, the order is described as *the* series of competencies rather than *a* series of competencies. There is an assumed correctness to the order in this choice of article (whether consciously intended or not), a hierarchy of learning through which one must progress in order to achieve the intended goal of giving the students a toolbox of decision heuristics. This assumption of a correctness of order is further confirmed in the explanation that the design is following the "'historic trajectory … from logic to probability" (p. 238). The strictness of defining and the authority of following the series of competencies given are representative of the valuing of rational and scientific knowledge and ways of knowing, with no consideration made of alternative kinds of knowledge or ways of knowing. The design of the activities also is grounded in rational and scientific knowledges and ways of knowing. For example, derivation of specific heuristics implies that these heuristics are most important, or singularly best, for considering risk and decision-making. These are not just any heuristics; however, they are heuristics that are bounded in rationality. In fact, rationality and scientific methods dominate this study's intended learnings, with categorizations of the kinds of activities with very specific and singular connections being made between them. Overall, Martignon and Krauss' study is well grounded within the kinds of knowledges and ways of knowing valued within the Traditional Western worldview, and in some ways, it supports the devaluation of other ways of knowing and kinds of knowledge, such as beliefs and emotional knowledge.

Valued Knowledge and Ways of Knowing in Risk Education – Ainley, Kapdia, Kent, Levinson, Pratt, and Yogui.

As noted earlier, the primary difference between the study done by Ainley, Kapdia, Kent, Levinson, Pratt, and Yogui and the study done by Martignon and Krauss is the former's acknowledgement and incorporation of affective or emotional responses and data. As such, in the areas of overlap, it will be assumed that the research of Kapadia, et. al. also values the ways of knowing and kinds of knowledge that are foundational to the Traditional Western worldview. Other kinds of knowledge and ways of knowing, however, can also be seen as valued within this research. First, the assumption held by the researchers regarding the importance of personal value systems and judgments (as noted by Kent et. al., 2010) while making decisions aligns with the valuing of emotional and possibly experiential knowledge and ways of knowing, which are valued within an Indigenous worldview. This assumption is further realized within the study proper through both the Painometer and Risk Mapping tools. The Painometer is explicitly about incorporating the emotional responses and knowledge of the participants into the study, and the Risk Mapping tool is an attempt to create a model detailing relationships between emotional responses and knowledge together with the rational and scientific knowledge captured within the Operation Outcomes. Even the specific design of the Operation Outcomes tool allows for emotional responses and knowledge to (implicitly) enter into the decision-making experience as the participants choose which outcomes to focus on while using the tool. One can easily assume that some, if not all, of these choices are subjectively determined based upon emotional and relationship-based knowledge. This emphasis on emotional and relationship-based knowledge and ways of knowing is representative of some of the kinds of knowledge and ways of knowing valued within an Indigenous worldview. Additionally, relationship-based knowledge and ways of knowing are also being encouraged by the design of the study, stepping beyond the typical boundaries (or compartments) of science and mathematics, inviting the mathematics teachers to consider the socio-scientific aspects of risk in the given dilemma, while encouraging the science teachers to consider the mathematical properties of probability at the same time. Finally, because the design of the study is such that no one kind of knowledge or way of knowing is overtly presumed to be of greater importance, the valuing of rational and scientific knowledge within the study can also be seen to align with an Indigenous worldview's consideration and acceptance of diverse ways of knowing and diverse knowledge depending upon the relationships sought and created within that context and at that time.

Unrecognized or Devalued Kinds of Knowledge and Ways of Knowing

As was stated previously, but warrants revisiting, when working within an ethical space, the above analyses and the ones to follow are not intended to serve as statements of judgments; rather, they are meant to give a different perspective, a new set of lenses through which risk, decision-making, and risk education can be contemplated and explored. For the same reasons as no identification of what kinds of knowledge and ways of knowing are being valued within different forms of communications about risk or within theories of risk and decision-making, the same two portions of the literature review will not be analyzed. Instead, the reader is invited to, at their leisure, reflect upon the possibilities from within their own knowledge and ethical spaces and experiences.

In reflecting upon the positioning of risk within curricular (standards) documents, the compartmentalization of risk into specific subjects is consistently present. As a result, the message (intended or not) conveyed is that risk, as mentioned in a particular place in a particular document, should be studied within the isolated conditions therein. As such, the possibility of building relationships within the learning is bounded to particular kinds of knowledge and ways of knowing. What is not being explicitly recognized (hence valued) is the many ways in which risk crosses the artificial boundaries of subject areas within everyday life. Situations of risk that one encounters in their personal life rarely are restricted to one field only. Financial risk is not just risky in relation to money, but to every aspect of life, just as what is involved in a health risk is never only medical in nature. Thus, there is an opportunity to investigate the possibility of risk education without specific subject boundaries. To an extent, Ainley et. al.'s study has done so, but the possibilities of relationship-based knowledge can extend further. Perhaps risk and decision-making could be considered as guiding principles behind education, as competencies, which are common to, and integrative of, all subject areas; risk literacy could be seen as foundational across all curricula

There are also other kinds of knowledge and ways of knowing that have yet to be valued (or acknowledged, even if to be dismissed) within risk education research including intuitional, spiritual, physical, experiential, traditional, and cultural knowledge. There is an opportunity for reflection upon, and possible inclusion of one or more of these kinds of knowledge and their associated ways of knowing within risk education, and such possibilities deserve candid consideration. Next, a small hint of what such considerations might include is given.

Why Consider the Unconsidered?

Some might dismiss, at this point, everything that has proceeded in this paper as over-exuberant and self-indulgent theorizing. Without any hint at why one might even consider any of the additional knowledges and ways of knowing mentioned, the whole idea can seem irrelevant and impossible to realize. For this reason, two real-life situations steeped in risk and decision-making are presented and reflected upon: the destructive tsunami of 2004, and the so-called "Navajo plague" of 1993. Each of these events took many lives, but each also reveals how some knowledges and ways of knowing that were ignored at the time may have changed the final outcomes.

Boxing Day, 2004

On Dec. 26, 2004, a magnitude 9.0 earthquake centered near the west coast of Sumatra and under the Indian Ocean occurred. The energy released by this earthquake has been estimated to be equivalent to 23 000 atomic equal to those of Hiroshima. The resulting tsunami a few hours later had waves that moved at the speed of a jet. (National Geographic News, January 5, 2005). Pictures and videos of homes, people, animals, and all kinds of belongings being swept into the ocean went viral across all forms of media. More than 150 000 people were killed, with some estimates being placed at 250 000 and higher, and millions lost everything they had. Despite all of the technology and scientific models available to predict and communicate risks of earthquakes and tsunamis, scientists were unable to provide adequate warning for the event. Yet, a number of indigenous groups, frequently ignored or seen as inferior to other non-indigenous inhabitants, survived en masse. Unfortunately, the same was not true for the non-Indigenous people inhabiting the same islands and communities. The mass survival of these indigenous peoples is now recognized to be the result of their "in-depth knowledge of the environment" (Mercer, Dominey-Howes, Kelman, & Lloyd, 2007, p. 251). As an example, the Moken (or sea gypsies), an Indigenous group from Thailand who live on the Indian islands of Andaman and Nicobar, "managed to anticipate the tsunami danger. Their knowledge of wind, tides, and the animals, which had been passed down from generation to generation, prepared them to deal with the natural disaster" (Perez, n. d., p. 1). Part of this knowledge included the silence of the cicadas, which was understood to tell the people to run for higher ground, and they did. When interviewers asked a Moken man why the tsunami had happened, he responded "The big wave had not eaten anyone for a long time, and it wanted to taste them again" (p. 2). Such knowledge would not, at least in Traditional Western worldview terms, be considered rational or scientific knowledge. Nor is it emotional knowledge. It is the traditional knowledge of the people, the knowledge which has been preserved and carried forward through generations of oral traditions. It is spiritual knowledge, intuitional knowledge, physical knowledge, and possibly experiential knowledge. No testing, isolation, compartmentalization, or abstraction of the knowledge was done. Similar examples from onslaught of the tsunami can be found throughout the region's Indigenous peoples, including those who live on Nias Island where, not only did the Indigenous people survive, but so too did their homes that were nearly 100 years old, while the new modern homes on the island were destroyed. These examples are filled with a kind of knowledge and a way of knowing that has been "increasingly recognized in the international arena, yet is frequently overlooked in practice" (Mercer, et. al., 2007, p. 247). Will the same be true in risk education?

The Navajo Plague

In the spring of 1993, a healthy, newly engaged, Navajo woman of 24 became sick one day with "a stuffy nose, a dry cough, aches, and little else. It looked like an ordinary case of the flu" (Arviso, & Cohen, 1999, p. 117). The following day, the woman "showed up in Crownpoint in severe respiratory distress and hypoxic ... She'd died a few hours later" (p. 118). On the day of her funeral, her 19 year old fiancé became similarly ill, was "brought to the GIMC emergency room in full respiratory and cardiac arrest and died shortly thereafter" (p. 120). These were the two patients of a soon to be epidemic that was spreading through the Four Corners – the name given to a region within Arizona, New Mexico, Colorado, and Utah in which a number of Navajo reservations are located. Although the as yet unidentified disease was seemingly targeting only Navajo people (hence the name 'Navajo Plague'), restaurants and businesses in communities adjoining the reservations began refusing to serve anyone who appeared to be Navajo in descent; people began cancelling vacation reservations in the south; and "the national media jumped to the conclusion that it was *because they were Navajo* that these individuals had contracted" (pp. 121-122) this acute respiratory distress syndrome. Local doctors and health care workers were dumbfounded as to the underlying cause for the disease, and so the Centre for Disease Control (CDC) was called in to solve the mystery. The CDC carried out a series of laboratory tests that "failed to identify any of the deaths as caused by a known disease such as bubonic plague" (Centers for Disease Control and Prevention, 2012). As additional testing continued, physicians and researchers repeatedly found that "The particular mixture of symptoms and clinical findings pointed… away from possible causes, such as exposure to a herbicide or a new type of influenza, and toward some type of virus". Tissue samples were analyzed by virologists at the CDC, ultimately leading to the identification of a previously undocumented type of hantavirus. The species of mouse (the deer mouse) known to carry and transmit this virus through its fecal matter and urine is not considered endemic to the Four Corners environment, and for this reason, hantavirus had not been considered in the original testing. Had hantavirus been included within the possible underlying diseases that were originally tested for, many people may not have died.

However, at least one of the healers within the Navajo reservations knew that a change in climate could result in the deer mouse being, at least temporarily, endemic to the Four Corners. In fact, a worker from the CDC, who was of Navajo decent had gone, to see the healer to ask about the disease. The healer replied by showing the CDC worker a photograph of a sand painting with a mouse in it, and he also told the worker that "many years ago such a sickness had occurred and that the sand painting had been used to treat it" (Arviso, & Cohen, 1999, p. 122). In reality, the sand painting did more than identify the particular breed of mouse responsible for the illness. It also explained why the population size of that particular kind of mouse would increase: three or more years of excessive rain leads to increased production of the seeds of the dwarf pine trees in the area, and those seeds are one of the best food sources for the deer mouse. When finishing his sharing of the story of the sand painting, the healer told the worker to share this knowledge with the CDC, and more specifically, to "'Look to the mouse'" (p. 122). Sadly, in all of the documented knowledge about this outbreak, there is no mention of anyone else (not even the Navajos living on the reservations) approaching the healer for information. Like the scientists, most Navajo people believed that this outbreak must be something new, beyond and foreign to their traditional knowledge. The CDC has officially acknowledged that: "Navajo Indians… recognize a similar disease in their medical traditions, and actually associate its occurrence with mice. As strikingly, Navajo medical beliefs concur with public health recommendations for preventing the disease" (Centers for Disease Control and Prevention, 2012). Perhaps, if the traditional knowledge and ways of knowing had been valued by the scientific practitioners and, dishearteningly, by the Navajo people themselves, the hantavirus diagnosis would have occurred sooner, and fewer young and promising lives would have been lost. Perhaps research into risk education would benefit from acknowledging the value of asking and listening to learn from traditional knowledge keepers.

In both the case of the 2004 tsunami and the 1993 hantavirus epidemic, the traditional knowledges of indigenous peoples, neither based in rational scientific knowledge and ways of knowing or in emotional and affective knowledge and ways of knowing, could have reduced the risk to the people in those situations. However, to acknolwedge and value such knowledge and ways of knowing, and to incorporate them into our understanding of risk, risk understanding, risk management, and risk-based decision-making, we need to come to terms with the limitations imposed by beliefs such as significant stochastic understanding being necessary for decision-making. Undoubtedly, there are times when such stochastic knowledge is of huge benefit in decision-making, but as the two previous examples demonstrate, stochastic knowledge (at least as understood within Western knowledge and mathematics) is not the only way in which one can make enlightened decisions. How this understanding will inform risk education remains to be seen.

Final Words

At this point, the relevance of including the discussion of research on communicating about risk and theories of risk and decision-making within the review of the literature can be elucidated. Specifically, the consideration of alternative kinds of knowledge and ways of knowing could also be considered within the areas of communicating about risk and theorizing about risk and decision-making. It is a discussion which might not only result in significant consequences to the research being done, but it also could help in the processes of decolonizing stereotypical and prejudicial ideas held about Indigenous peoples and their knowledge as well as the resulting oppression and marginalization.

Most of the research related to risk education and decision-making has been, for the most part, unaware (or possibly dismissive) of many of the ways of knowing and kinds of knowledge valued within an Indigenous worldview. Perhaps this is due to the Traditional Western worldview's (which arguably has grounded much of academic research) unquestioning belief in absolute truth and rationality; perhaps it is due to the need to create the 'other' to find self-worth (another consequence of the underpinnings of the Traditional Western worldview and processes of colonization and oppression); or, perhaps it is due to the failure of an education system and its supporting research to include considerations of alternative ways of knowing and kinds of knowledge. The reason why is not of greatest importance; rather what is important is what will be done about this realization.

The emerging field of risk education has within its reach an opportunity that has been missed or ignored by so many other fields of study: the valuing of alternative ways of knowing and kinds of knowledge beyond those that are rationally and scientifically-based and the emerging valuing of emotions and emotional responses. Perhaps this time, it will not be that "western science remains firmly entrenched in its traditional methods. An entrenchment from colonial times… when our ignorance of the 'other' contributed to an increased divide between them (the developing world) and us (the Western world)" (Mercer, et. al, 2007, p. 246). Perhaps risk education will help break through these, often unspoken, boundaries.

As a final note, if the decision is made to embark upon the consideration and possible inclusion of other kinds of knowledges and ways of knowing, care must be taken to not appropriate and commodify those kinds of knowledges and ways of knowing (Smith, 2000). In the valuing of new kinds of knowledges and ways of knowing, then so to must the keepers of those knowledges and ways of knowing be acknowledged, valued and respected. Perhaps, risk education will not only venture into new worlds of knowing — it is possible that it could contribute to the processes of decolonization for the good of all peoples.

Acknowledgements

In keeping with the sentiments expressed in the final words above, I would like to acknowledge the contributions of the Indigenous scholars, elders, knowledge keepers, and ancestors, who despite the challenges of colonization, have managed to preserve their worldviews. Further, I thank them for sharing their worldviews with everyone, despite the condemnation and oppression they have so often experienced as a result of holding such worldviews.

References

Absolon, K. (2010). Indigenous wholistic theory: A knowledge set for practice. *First Peoples Child & Family Review, 5*(2), 74–87.

Absolon, K., & Willett, C. (2005). Putting ourselves forward: Location in Aboriginal research. In L. Brown & S. Strega (Eds.), *Research as resistance: Critical, Indigenous, & anti-oppressive approaches*. Toronto, ON: Canadian Scholars' Press.

Arviso Alvord, Dr. L., & Cohen Van Peet, E. (1999). *The Scalpel and the Silver Bear.* Toronto, ON: Bantam Books.

Barnhardt, R., & Kawagly, A. O. (2005). Indigenous knowledge systems and Alaska Native ways of knowing. *Anthropology and Education Quarterly, 36*(1), 8–23.

Borovcnik, M., & Kapadia, R. (2011). Determinants of decision-making in risky situations. *International Statistical Institute: Proceedings of the 58th World Statistical Congress (pp. 5503-5508).* Dublin, Ireland.

California Department of Education (2009). *Health Education Content Standards for California Public Schools, Kindergarten Through Grade Twelve.* California Department of Education. Online: http://www.cde.ca.gov/be/st/ss/documents/healthstandmar08.pdf.

California State Board of Education (2014). *Content Standards.* Online: http://www.cde.ca.gov/be/st/ss/index.asp

Canadian Council on Learning (2007). Redefining how success is measures in First Nations, Inuit and Métis Learning. Retried from: http://www.ccl-cca.ca/pdfs/RedefiningSuccess/Redefining_How_Success_Is_Measured_EN.pdf.

Centers for Disease Control and Prevention (2015). *Tracking a Mystery Disease: The Detailed Story of Hantavirus Pulmonary Syndrome (HPS).* Online: http://www.cdc.gov/hantavirus/hps/history.html.

Common Core State Standards Initiative (2015). *High School: Statistics & Probability >>Using Probability to Make Decisions.* Online: http://www.corestandards.org/Math/Content/HSS/MD/#CCSS.Math.Content.HSS.MD.B.5.

Department of Education (2013, September). Citizenship programs of study: key stages 3 and 4. National Curriculum in England. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239060/SECONDARY_national_curriculum_-_Citizenship.pdf.

Department of Education (Sept. 11, 2013). Guidance: Personal, Social, Health and Economic Education. Online: https://www.gov.uk/government/publications/personal-social-health-and-economic-education-pshe/personal-social-health-and-economic-pshe-education#contents.

Department of Education (2013, September). Science programmes of study: key stage 3. National Curriculum in England. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/335174/SECONDARY_national_curriculum_-_Science_220714.pdf.

Department of Education (2014, December). Science programmes of study: key stage 4. National Curriculum in England. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/381380/Science_KS4_PoS_7_November_2014.pdf.

Department of Education (2014, July). Mathematics programs of study: key stage 4. National Curriculum in England. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331882/KS4_maths_PoS_FINAL_170714.pdf

Ermine,W. (1995). Aboriginal epistemology. In M. Battiste & J. Barman (Eds.), *First Nations Education in Canada: The Circle Unfolds* (pp. 101–112). Vancouver: UBC Press.

Ermine, W. (2007). The ethical space of engagement. *Indigenous Law Journal, 6*(1), 193–203.

European Centre for Disease Prevention and Control (January 30, 2015). *Outbreak of Ebola Virus Disease in West Africa*. Ninth update, 30 January, 2015. Stockholm: ECDC. Online: http://www.ecdc.europa.eu/en/publications/Publications/RRA-Ebola-Feb-2014.pdf

Gal, I. (2012). Developing probability literacy: Needs and pressures stemming from frameworks of adult competencies and mathematics curricula. *Proceedings of the 12th International Congress on Mathematical Education*, Seoul, Korea, 1-7. Online: http://www.icme12.org/upload/upfile2/tsg/2088.pdf.

Hogan, L. (2000). A different yield. In M. Battiste (Ed.), *Reclaiming Indigenous Voice and Vision* (pp. 115– 123). Vancouver: UBC Press.

Kappo, T (Dec. 19, 2014). *Harper's comments on missing, murdered aboriginal women show 'lack of respect'.* CBC News. Online: http://www.cbc.ca/news/aboriginal/stephen-harper-s-comments-on-missing-murdered-aboriginal-women-show-lack-of-respect-1.2879154.

Kent, P., Pratt., D., Levinson, R., Yogui, C., & Kapadia, R. (2010). Teaching uncertainty and risk in mathematics and science. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8)*, Ljubljana, Slovenia.

Kovach, M. (2009). *Indigenous methodologies: Characteristics, conversations, and contexts.* Toronto: University of Toronto Press.

Leavitt, R. (1995). Language and cultural content in Native education. In M. Battiste & J. Barman (Eds.), *First Nations Education in Canada: The Circle Unfolds* (pp. 124–138). Vancouver: UBC Press.

Little Bear, L. (2000). Jagged worldviews colliding. In M. Battiste (Ed.), *Reclaiming Indigenous voice and vision* (pp. 77–85). Vancouver: UBC Press.

Macilwain, C. (2014). A touch of the random. *Science*, *344*(6189), 1221-1223.

Martignon, L., & Krauss, S. (2009). Hands-on activities for fourth graders: A tool box for decision-making and reckoning with risk. *International Electronic Journal of Mathematics Education, 4*(5), 227-258.

Mercer, J., Dominey-Howes, D., Kelman, I., & Lloyd, K. (2007). The potential for combining indigenous and western knowledge in reducing vulnerability to environmental hazards in small island developing states. *Environmental Hazards: Human and Policy Dimensions, 7*, 245-256.

Meyer, M. A. (2003). *Ho'oulu: Our time of Becoming.* Honolulu: 'Ai Pōhake Press.

National Geographic News (January 7, 2005). *The Deadliest Tsunami in History?* Online: http://news.nationalgeographic.com/news/2004/12/1227_041226_tsunami.html.

Perez, F. Y. L. (n.d.) *Survival Tactics of Indigenous Peoples.* Online: http://academic.evergreen.edu/g/grossmaz/LEEPERFY/.

Persky, J. (1995). Retrospectives: The ethology of *Homo Economicus*, *Journal of Economic Perspectives, 9*(2), 221-231.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning, 13*(4), 322-345.

Pratt, D., Levinson, R., Kent, P., & Yogui, C. (2011, February). *Risk-based decision-making by Mathematics and Science Teachers.* Paper for the CRME-7 Conference, Rzeszów, Poland. Online: https://www.cerme7.univ.rzeszow.pl/WG/5/CERME_Pratt-Levinson-Kent-Yogui.pdf

Pratt, D., & Yogui, C. (2010, August). *A Constructionist Approach to a Contested Area of Knowledge.* Paper presented at the Constructionism 2010 Conference, Paris, France.

Public Health England (February, 11, 2015). *Risk Assessment of the Ebola Virus Disease outbreak in West Africa.* PHE publications gateway number: 2014578 Published (v5): 11 February 2015. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/403092/Ebola_Risk_Assessment_Update_v5.pdf

Russell, G. L., & Chernoff, E. J. (2013a). Incidents of Intrusion: Disruptions of Mathematics Teaching and Learning by the Traditional Western Worldview. *Proceedings of the 35th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Chicago, IL, USA.

Russell, G. L., & Chernoff, E. J. (2013b). The Marginalization of Indigenous Students within School Mathematics and the Math Wars: Seeking Resolutions within Ethical Spaces. *Mathematics Education Research Journal of Australia, 25*(1), 109-127.

Russell, G. L., & Chernoff, E. J. (2012). Unifying challenges in the teaching and learning of mathematics: Two can become one. *Proceedings of the 34th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Kalamazoo, MI, USA.

Russell, G. L., & Chernoff, E. J. (2011). Transforming mathematics education – applying new ideas or commodifying cultural knowledge? *33rd Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Reno, NV: University of Nevada.

Saskatchewan Ministry of Education (2013a). Foundations of Mathematics 30. Regina: Saskatchewan Ministry of Education.

Saskatchewan Ministry of Education (2013b). Workplace and Apprenticeship Mathematics 30. Regina: Saskatchewan Ministry of Education.

Schelbert, L. (2003). Pathways of human understanding: An inquiry into Western and North American Indian worldview structures. *American Indian Culture and Research Journal*, 27(1), 61–75.

Schweingruber, H., Keller, T., & Quinn, H. (Eds.). (2012). *A Framework for K-12 Science* Education*: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.

Simon, H. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review, 49*(3), 253-283.

Smith, G. H. (2000). Protecting and respecting Indigenous knowledge. In M. Battiste (Ed.), *Reclaiming Indigenous voice and vision* (pp. 209-224). Vancouver: UBC Press.

Spiegelhalter, D. (2014). The future lies in uncertainty. *Science*, *345*(6194), 264-265.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science, 333*, 1393-1399.

Till, C. (2014). Fostering risk literacy in elementary school. *Mathematics Education, 9*(2), 83-96.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. Science, 185(4157), 1124-1131.

WNCP. (2008) Common Curriculum Framework for Grades 10 -12 Mathematics. Western and Northern Canadian Protocol. Online: http://wncp.ca/media/38771/math10to12.pdf.

Youngblood Henderson, J. (2000). Postcolonial ghost dancing: diagnosing European colonialism. In M. Battiste (Ed.), *Reclaiming Indigenous Voice and Vision* (pp. 57–76). Vancouver: UBC Press.

Risks Worth Taking?
Social Risks and the Mathematics Teacher

Ami Mamolo
University of Ontario Institute of Technology, Canada
Laura Elizabeth Pinto
University of Ontario Institute of Technology, Canada

Abstract: In this article, we explore notions of risk as perceived or experienced by individuals involved in mathematical education. We present this exploration in the form of vignettes, each illustrating a form of risk: a parent's reaction to classroom "propaganda"; a teacher trying to do justice by her students; a teacher confronted by his administration; and a college professor who believes university policy to be unjust. Each vignette sheds light on areas in which teacher education may offer additional support in fostering the mathematical knowledge, pedagogical sensitivity, and social awareness required to foster, what are in our view, much needed risks in the mathematical (and otherwise) education of pupils. Following the vignettes, we offer a discussion of factors that contributed to the risks perceived or experienced by teachers: neoliberal discourses, and the powerful cultural scripts that leave teachers feeling that they must hold all control, authority, and knowledge.

Keywords: risk, mathematics education, social justice, liberation, teacher development.

Introduction

The project of education is fraught with immense vulnerabilities faced by both teacher and student, yet the nature of risk in education is different from more calculable risks in other fields of endeavour. Notions of *risk* – mathematical or otherwise – are broadly interpreted in the literature. Risks manifest themselves within mathematical activity, social activity, and within interactions of the two. Risk necessarily refers to a future event, to one or more behavioural choices, the probability of which is amenable to calculation (Pellizzoni, 2011).

In our perspective, mathematical risk extends beyond probability theory – and includes risks associated with expanding, applying, or re-imagining one's mathematical understanding and the potential of mathematics to yield understandings about broader, complex, and social issues. Indeed, recent research has shed light on the power of mathematics in preparing students to understand and address issues of social injustice (e.g., Gutstein, 2003, 2006; McCoy, 2008). Teaching mathematics for social justice can offer students "instruction that includes the mathematics deemed necessary for success in the current [school] system while simultaneously providing students an opportunity to use mathematics to expose and confront obstacles [in society]" (Bartell, 2011/2013, p.1). Critics of current school systems have pointed to the responsibility of "teachers and the education system to extend possibilities for human development" noting the difficulties of fulfilling such responsibility given the tendency to pay "more attention on test results than personal development" (Watson, 2006, p.3).

Researchers point toward the role of mathematics in empowering those who are marginalized as it can offer a means through which to interpret social institutions, traditions, and potential political reform (Skovsmose, 1994). Gutstein (2006) writes of the importance of using mathematics to change the world, as it can foster in students a sense that they are "capable of making change" and may help

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 85-94

them develop "a sense of social agency" (p.27). In line with these perspectives, Bartell suggests, "The purpose of education is not to integrate those who are marginalized into the existing society but rather to change society so that all are included" (2011/2013, p.3). While such orientations toward mathematics education resonate within academic communities, they appear contentious in the public sphere where traditional views of "teacher" and an expectation of the austere neutrality of mathematics still abound. Promoting critical perspectives on social issues is positioned as "brainwashing" and the idea that children might take action to initiate social change is viewed as "politically motivated" and "risky" (Reynolds, 2012). Such perspectives are largely unsubstantiated, though we agree that confronting them involves risk.

A socially just view of schooling emphasizes the educational aim of "emancipation from traditional custodianships and intellectual sensibility" and is positioned as "a pathway to human flourishing, both personal and social," and as such is fraught with "inescapable risks" (Smyers & Hogan, 2005, p.119). At stake is what *"students become as human beings* as a consequence of what [they] experience as learners" (Smyers & Hogan, 2005, p.115, emphasis in original). In particular, the transformative process of becoming a mathematics teacher, in our view, requires a liberation from restrictive, but popular, conceptions of what is, and for what is, (school) mathematics, and includes the inherent and multifaceted risks faced by the individual committed to the activation of new perspectives on familiar ideas.

On all counts, these social risks involve trust, autonomy, and vulnerability where teachers and learners experience tensions between and amongst individuals and ideas as they are shaped and structured by the systems within which mathematics education exists. Though calculable uncertainty creates structure and organization in a complex world, non-calculable uncertainty prevents humans from being prisoners of an inevitable path: the future and innovation cannot be predicted by that which is statistically knowable (Pellizzoni, 2011). Bayesian theories of subjective probability refer all probabilities to the agent's knowledge, "because relative frequencies are only sample data of past events that influence subjective probabilities of future events" (Pellizzoni, 2011, p.797). Building on this perspective, we view the subjective risks associated with the intersection of mathematical and social activity as relative to an individual's perspective and knowledge. Such risks are not calculable. They are a projection of sample data from past events – risks experienced in the past and their corresponding consequences – onto the possibilities for future events. Thus, while an individual may "weigh" the perceived risks inherent in a particular course of action, it is their subjective interpretation of past events which influences their expectations for the future. For us, this raises an important question: to what degree should past events (in mathematics education) determine which risks are worth taking? Though we do not attempt to answer this question, its consideration informed our approach in this article.

Through a collection of vignettes, we explore notions of risks that were perceived or experienced by mathematics educators attempting to address issues of social justice in their instructional practice, and we consider factors that may have contributed to these perceptions or experiences. Our vignette's present an *account-of* risks, while our discussion provides an *account-for* them. Account-of refers to descriptions which avoid as much as possible evaluation, interpretation, or explanation, while an account-for interprets, analyses, and explains (Mason, 2002). We conclude with a discussion of key themes arising from the vignettes. One of these is the way in which neoliberal discourses shape false "value neutrality" leading to immense risk-taking when social justice issues – perceived as "biased" – are taken up by teachers. The second point of discussion is the way in which traditional and conventional approaches to mathematics learning rely on cultural myths of "teacher control," leading to risks when disciplinary knowledge is expanded to take on social justice. In our discussion, we pose important questions for teachers and teacher educators to consider about both current practice, and the potential to transform mathematics learning by taking risks with curriculum and pedagogy.

Vignettes

Taking Risks in the Syllabus

Bob Humphries took a deep breath before he knocked on the Department Chair's door. The email requesting the meeting didn't specify its purpose, but Bob was almost certain he knew what it was about. Syllabi had just been submitted for review; Bob had taken a big risk now that he had achieved tenure.

"Have a seat, Bob," Tony said as he gestured towards the conference table. "I wanted to talk to you about that paragraph at the end of your syllabus."

"Tony, since the day I started teaching here, I've been vocal about these concerns at just about every faculty meeting. Nothing in that syllabus or that paragraph is untrue – in fact, if anything it's an invitation to get our students to engage in some research and critical thought about the implications of grading on a curve" (for example, Hout, Jankowski, Lucas, Swidler & Voss, 1996). Tony glanced at paragraph in question again:

> The instructor assigns grades based on evidence of your mastery of the skills and ideas on which the course turns. He considers the grading policy promulgated by the Department of Mathematics and Statistics which requires him to adjust grades on the curve to be manifestly unjust and to verge on irrationality. He will definitely not lower any student's grade in order to conform to that policy. He thinks that if you are well-to-do, you should consult a lawyer wherever you have concrete reason to believe that application of such arbitrary policies has deprived you of grades to which your performance entitles you. If you are not well-to-do, you are well acquainted with injustice and have, of necessity, learned to live with it.

"But the message, Bob, you know it's not quite what we hope to convey to our students. It's a well-accepted institutional practice, and one that's not unique to this university."

"I'll say it again, Tony: there's nothing false in that syllabus, and if I'm not mistaken, I can exercise my academic freedom in drafting the content of any and all syllabi."

"You know I respect you Bob, and I'm not going to argue it. But I have to submit these to the Dean, and I wanted to give you a chance to make revisions, if you want, before she sees it."

"Send it along as is, please, Tony."

Bob stood at the copier 3 days later, making a duplicate of the folder he'd labeled "Bell Curve" in preparation for his meeting with the Dean. The dossier was filled with references that supported his position on the syllabus controversy. The meeting mirrored the conversation he had with Tony, and the Dean could not argue with Bob's position.

"Ultimately," she said at the conclusion of the meeting, "it's your right to leave it in, Bob. But I really wish you'd take it out. It's opening the Faculty up to all sorts of objections and problems."

Who Run this Motha?

Many of the girls in Alex's grade 7 class had taken to singing Beyoncé's hit "Run the World (Girls)" (Nash, Knowles, van de Wall, Pentz, Tayor & Palmer, 2011) at recess. "Who run the world?" one chanted; "Girls!" the others replied.

"Who run this motha?"

"Girls!" (Nash et al., 2011)

While he encouraged positive messages of empowerment, he wondered if Beyoncé's brand of "girl power" discourses helped to motivate, or concealed realities that needed to be openly discussed if any real change was to take place (see, for example, Pinto, 2011).

"Ready to do some math?" he enthusiastically asked the class one morning. He'd prepared a lesson that invited students to investigate wage inequity and the reasons behind it. He created a "Web Scavenger Hunt" with clues and links, which led the class to various current and reliable sources that outlined wage gaps. He also created simplified online visuals to help students visualize Charles and Grusky's (2007) gendered "occupational ghetto" concepts, expose the ways in which women continue to face both segregation within occupations on two axes (the horizontal axis representing occupational types, and the vertical axis, representing hierarchical representation). The Web Scavenger Hunt culminated in students graphing wage differences and creating a display. This was a jumping off point for a social studies lesson that explored the reasons behind, and controversies surrounding, the data.

Alex was surprised to learn that an angry parent had called to complain about the "propaganda" in class, and sent an article over (Lukas, 2012). The school principal called a meeting with the parent. "Alex, remember we have an obligation to the parents. They are our customers, and we have to make sure our classes reflect community values."

Customers!? Alex thought. "Look, I know we have to be tactful at the meeting, but the fact is the students went on an inquiry-based learning project, and used real data to solve a math question. This is good teaching. Authentic! Dare I say evidence-based? I don't see why we're even entertaining this complaint with a meeting. In a democracy, don't we need to consider perspectives like this, even if we don't agree? After all, don't we all have an obligation to support this school board's policy that encourages the teaching of controversial issues?"

Alex sat in the office with the principal and the student's father to sort things out. The irony of three men discussing the issue at hand was not lost on him. He could hear the girls from his class sing, "Who run this motha?" Definitely not girls.

In a Position to Take a Risk

Sonya was in her first year of teaching. She was meeting with Paul, who had been teaching at her school for the past 15 years. They were discussing a math project Paul recently gave to his class. The project was an inquiry-based exploration of the accessibility of healthy foods in poverty-stricken communities. Paul described the project:

> "We had the opportunity to visit the local food bank. Before we went there, we went to a supermarket to purchase some non-perishable items that we could donate. I gave students a budget of $2.00 each and watched while they navigated the aisles looking for items. After a while, students started teaming up to purchase larger items, and they started critiquing name-brand items versus discount items. They were super pumped when one of them found a good sale and shared it with the rest. When we got to the food bank, students had the opportunity to ask some of their questions, and then in class we unpacked everything. We focused on place value and numbers, but I think there is a good data management connection, so I'm going to try to bring them [students] back to these questions when we get to that unit."

Sonya interjected: "I understand that teaching is not all about the curriculum, and that it's possible, even necessary sometimes to take math outside of the classroom and into the real world, but how did you deal with students' questions about poverty? I mean, they have nothing to do with the math curriculum, so like wouldn't this make more sense in a social studies class?"

"Well, yes you do need to keep in mind the curriculum, but you also need to be able to move along the path with your students and not close any doors, " Paul responded.

Sonya was unconvinced. "OK, I do believe that it's important to help students see how math can connect to issues in society, but to invest the amount of time that this project took… aren't you jeopardizing the students' learning? It just seems unfair to students to not spend more time on the math concepts outlined in the curriculum."

As their conversations continued, Sonya softened her position and said she might consider doing such a project with her own class – but after some adaptation. Paul was happy to collaborate in re-designing the lesson, and the conversation ended there.

Later, on reflection, Paul mused:

> "I think it takes a great leap of faith to become comfortable with the uncomfortable. What's motivated me over the past five years is my graduate work. I learned about many different modalities of teaching, different philosophies about math and math learning and different ways of thinking about particular math topics. This has definitely helped me justify my teaching methods, and has maybe made my approach feel a little more 'comfortable'."

What Math is All About

Leslie was teaching an integrated math and science unit on ecosystems where students worked in groups, moving through different learning centres. There were resources and guiding questions at each centre, but essentially the students were able to discover and research about almost anything that they wanted in relation to the topic of that particular centre. During one work period, Leslie's principal came to observe the class. They had what Leslie considered a "weird" exchange. The principal was curious as to whether Leslie was going to incorporate what was currently happening in the world – there had just been a major natural disaster, a tsunami, and the principal thought it would make a great current events segment.

Leslie happily informed the principal that there was one eco-centre dedicated to global climate change and natural disasters and that the students had a choice as to what they would like to study. Leslie sensed the principal wasn't too pleased with her response. She then listened as he proceeded to tell her how science and math are important:

> "Science is only good to teach one thing, procedural writing, and that's why we've taken that part out of the literacy program. Most of these kids are going to end up in workplace math, so what they need to know are the basics. They need to know about percents and finance, things that will actually be relevant to their lives. Generally, math is a study which contains questions and definite answers. There's usually one right answer and many wrong answers. Having a debate or discussion on such a topic [climate change] is not suited for any math course. There are more efficient and less debatable ways to explore math concepts. The last thing a teacher would want is to be accused of spreading propaganda."

Later, Leslie reflected on this exchange:

> "I've experienced a lot of negativity from colleagues, including administrators, along the lines of not really understanding what I'm trying to accomplish and that I couldn't possibly be teaching the students as I'm not standing in front of the class and that the students are relatively 'noisy'. It seems like a great number of teachers are quite uncomfortable straying too far from the prescribed curriculum or textbook, lesson, long-range plans. In reality, taking risks is part of the job, but also feeling that you have support from your administration in taking those risks is paramount. Having a supportive base or foundation to take these risks is important."

Discussion

In our discussion of the vignettes, we account-for the risks involved by interpreting them via two separate, but complementary, lenses: the imposition of "value-neutrality" and a broadened view of mathematical knowledge for teaching.

Neoliberal Discourses that Promulgate "Value Neutrality"

A common theme in the vignettes just presented is the tension between a desire to "do" social justice in mathematics education, and a belief that schooling is or should be "value neutral," which includes avoiding controversial topics. In Vignette 1, Bob wanted to call attention to injustice by way of a seemingly neutral grading policy – something university administrators do not want brought to the fore. In Vignette 2, Alex attempted to infuse a social justice mathematics project to specifically address the interests of his students, and invited them to unpack "girl power" discourses. He faced opposition from a "customer" and was not supported by his principal. In Vignette 3, Sonya had the impression that addressing social justice issues would jeopardize learning about mathematical concepts – suggesting that somehow mathematics needed to be divorced from social issues to maintain an air of neutrality. In Vignette 4, the principal suggested that incorporating climate change in math or science might lead to an accusation of "spreading propaganda" – an indirect implied assumption about avoiding controversy and adhering to something like value neutrality. In all cases, teachers took on perceived risks in order to infuse social justice into students' mathematics experiences. And in all cases, the substantive nature of social justice issues was perceived as controversial.

Views like those expressed in the vignettes – that education is or ought to be value-neutral – have "made something of a comeback in recent years" (Roberts 1998, p. 30, in Pinto, 2012). Value-neutrality is alluring in that it suggests one can avoid making choices, thus standing above or avoiding controversy. In truth, all educational choices involve values. Deciding what is "worth knowing" or "most important" are value-laden acts (Pinto, 2012). Mistaking value neutrality as a characteristic of education undermines the goal of social justice. Items that appear in curriculum documents privilege certain knowledge, skills, and attitudes. For instance, in mathematics, "axiomatic systems are greatly valued over less systematic forms of deduction including problem solving, calculation and unsystematised proofs" (Ernest, 2007, p.7), yet "such values have only been prominent during a small part of the history of mathematics" (p.4). In particular within mathematics, values underpin the conventions, approaches, and nature of what are viewed as acceptable ways of engaging in the discipline (Ernest, 1998). Why, then, are some packets of knowledge, or skills, or attitudes, viewed as value-neutral or controversial? Part of the answer seems to lie in the pervasiveness of neoliberal discourses that have dominated education policy and practice since the 1990s (Pinto, 2012; Wright, 2012).

Neoliberal positions are portrayed as neutral, while opposing positions are constructed as 'political,' and contrary to common sense (Pinto, 2012; Roberts, 1998). Social justice is perceived by the public (including parents) to have "a strongly progressive bent, and the idea of political manipulation creates fiercely negative reactions among parents" (Reynolds, 2012). Reynolds (2012) offered the example of a PETA poster angering a parent as a social justice initiative for bringing politics into the classroom. Yet, any decoration – whether poster, image, or anything else – would be equally value-laden. Some values, however, become "invisible" and perceived as unbiased (such as the salience of axiomatic systems in mathematics), while others (such as PETA's stance) are labeled as biased, or even, according to the principal in Vignette 4, propaganda. The expectation of objectivity or objectivism as inherent to mathematics lends itself well to a value-neutral standpoint and precludes the socio-cultural nature of mathematics teaching and learning.

Considering the vignettes through a value-neutrality lens raises a number of interesting questions for teachers and teacher educators. We suggest that it is important to raise such questions for discussion during teacher preparation (as well as in our own research). By raising the awareness of value-neutrality traps, teacher educators may support prospective teachers in articulating the choices they might make in instructional situations that extend views of mathematics beyond traditional expectations. In our own teaching, we have raised for discussion questions such as: What "values" exist in your school that are so engrained they become invisible? In textbooks or other curriculum artefacts you regularly use, whose values, perspectives, and positions are reflected? Whose are absent? Why or why not? Thinking critically about the curriculum policy that impacts your work, what underlying assumptions exist that are invisible? Whose perspectives or views about mathematics are privileged in policy? What risks does a teacher take when he or she takes a social justice approach in a mathematics classroom? What systemic factors (policy, parents, administration) exist? What factors might make a teacher more or less likely take social justice risks?

Broadening Disciplinary Inquiry and Knowledge

Gutstein's (2006) model of Teaching Mathematics for Social Justice identified changing the learner's orientation towards mathematics as one of its central aims. He described the importance of helping students transition "from seeing it [mathematics] as a series of disconnected, rote rules to be memorized and regurgitated, to a powerful and relevant tool for understanding complicated, real-world phenomena" (p.30). Similarly, teaching mathematics *through* social justice (that is, contextualizing mathematics within issues of social justice) can require and foster a similar shift in orientation. However, this also requires a shift in orientation towards mathematics *teaching* (Mamolo, 2014) -that is, teachers must emancipate themselves from the view that mathematics is either "right" or "wrong", that there is one privileged solution or approach, that the end result of a mathematics problem should be neat, tidy, and absolute. In the context of the real world, and especially with respect to issues of social justice, ambiguities and values cannot be avoided.

Watson (2006) emphasizes "the role of teachers and the education system to extend possibilities for human development" (p.3). In our view, enacting such an orientation requires breaking away from traditional views of teacher as the authority, in control of the knowledge to be learnt by students. Breaking away from tradition, as our vignettes suggested, is laden with inherent risks. For teachers to perceive that risk-taking is possible and worthwhile requires a position in which they have either the appropriate standing within their workplace to do so, or a knowledge-base which allows them to justify such teaching decisions. We interpret elements of both in our vignettes, but for the purpose of this discussion focus on the latter.

Teachers in other research avoided taking risks in the classroom when they perceived themselves as lacking sufficient subject-matter knowledge, and so they relied heavily on textbooks as a professional crutch, and for intellectual support (Pinto, McDonough, & Boyd, 2011). Given that textbooks tend to under-emphasize social justice in their "conventional" treatment of subject matter (e.g., Pinto, 2007), reliance on them works against social justice aims. Teachers' confidence in challenging publicly what they themselves are only just learning may be low for fear of undermining their classroom authority (Nuthall, 2004; Pinto, McDonough, & Boyd, 2011). For example, when asked why he felt the need to conform closely to the textbook, a high school teacher Philip explained, "Well it's basically because of my inexperience, I guess you could say, like, if I was probably more confident with the subject matter and the curriculum" (Pinto, McDonough, & Boyd, 2011).

Paul's musing in Vignette 3 reflected a similar sentiment in that his more extensive experience with mathematics created a 'broader' comfort zone and as such minimized certain risks. Sonya's concern that students' questions about poverty would have nothing to do with mathematics resonated with the difficulties prospective teachers can have in noticing relevant mathematical concepts when contextualized in issues of social justice (e.g., Mamolo & Martin, 2013). Similarly, the principal in Vignette 4 and the parent in Vignette 2 both conveyed very particular (limited) perceptions of what mathematics (and consequently mathematics learning) "should look like." Conventional and enduring "cultural myths" (Nuthall, 2004) place the teacher in control (at the front of the class with quiet students), the curriculum and textbook as the authorities of what and how students should be taught, and undebatable truths as the requisite knowledge to be acquired. There are risks involved on all three counts – social risks involved in managing a class that seems "out of control", subject-matter risks in diverting from prescribed approaches, and an intersection of social and subject-matter risks in negotiating disparate interpretations of mathematical "truths."

In accordance with Mason and Davis (2013), we interpret teachers' disciplinary knowledge of mathematics as a mode of being which influences how learning situations are structured, how student actions are interpreted, the flexibility with which teachers respond to such actions, and the in-the-moment decisions teachers make to foster and support student success. So again we ask: What factors might make a teacher more or less likely to take a social justice risk? What ways of being with mathematical knowledge can support or hinder such risk-taking? How can teacher educators support such ways of being?

Conclusion

Ernest (2007) argued that "it is no longer enough to claim that [values] are outside of its proper subject matter" (p. 11). The four vignettes presented here point to the ways in which teachers experience risk when attempting to address social justice issues in mathematics classrooms. Prior research as well as anecdotal experience (especially Reynolds, 2012) suggest that some teachers shy away from more topics out of fear of antagonizing parents – a that fear is rooted in avoiding risk by estimating noncalculable probability.

We recognize that taking risks in the classroom is dangerous business – especially in light of contemporary, prescriptive neoliberal policy that governs classroom practice, and the accompanying accountability apparatus that attempts to enforce compliance. Yet, the compelling goal of helping students flourish (Smyers & Hogan, 2005), in our view, justifies the risks associated with transformative mathematics education. To that end, this paper offered four examples of ways in which mathematics teachers took on risks associated with social justice. We identified some of the systemic and structural barriers that they faced. We also posed a number of questions in our discussion that we hope mathematics teachers and teacher educators will take up. These questions offer an invitation to reflect on the opportunities and strategies that may move teachers and teacher educators towards transformative mathematics education.

Our paper raises a number of issues that warrant future research. We believe that an agenda of narrative or case study research providing examples of successful transformative mathematics education – which would go beyond merely instructional descriptions to include the risks taken – would provide a useful basis for professional learning and transformative practice. As well, additional research detailing students' perspectives on engaging in social justice mathematics education would shed light on how such undertakings affect students and their learning. The research suggestions just described might also be beneficial in informing policy production at the district, province/state, and national levels.

Acknowledgements

This research was funded in part by the Social Sciences and Humanities Research Council of Canada.

References

Bartell, T. (2011/2013). Learning to teach mathematics for social justice: Negotiating social justice and mathematical goals. *Journal of Research in Mathematics Education, 41*, 5-35.

Charles, M., & Grusky, D. B. (2007). *Egalitarianism and gender inequality. The Inequality Reader: Contemporary and Foundational Readings in Race, Class, and Gender*, (p.327-42). Boulder, CO: Westview Press.

Ernest, P. (2007). The philosophy of mathematics, values, and keralese mathematics. *The Montana Mathematics Enthusiast, 4*(2), 1-13.

Ernest, P. (1998). *Social Constructivism as a Philosophy of Mathematics*. Albany, New York: SUNY Press.

Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education, 34*(1), 37–73.

Gutstein, E. (2006). *Reading and writing the world with mathematics: Toward a pedagogy of social justice.* New York, NY: Routledge.

Hout, M., Jankowski, M. S., Lucas, S. R., Swidler, A., & Voss, K. (1996). *Inequality by design: Cracking the bell curve myth* (p. 152). Princeton, NJ: Princeton University Press.

Lukas, C. (2012, April 4). It's time that we end the equal pay myth. *Forbes* [online edition]. Retrieved from: http://www.forbes.com/sites/realspin/2012/04/16/its-time-that-we-end-the-equal-pay-myth/

Mamolo, A. (2014). Noticing the math in issues of social justice. *Proceedings of the* 17*th SIGMAA on RUME Conference.* Denver, USA.

Mamolo, A., & Martin, L. (2013). Mathematical understanding in a social justice context. *Proceedings of the 34th International conference for Psychology of Mathematics Education – North American Chapter.* Chicago, USA.

Mason. J. (2002). *Researching Your Own Practice: The Discipline of Noticing*. London: Routledge Falmer.

Mason, J., & Davis, B. (2013). The importance of teachers' mathematical awareness for in-the-moment pedagogy. *Canadian Journal of Science, Mathematics and Technology Education*, *13*(2), 182–197.

McCoy, L. (2008). Poverty: Teaching mathematics and social justice. *Mathematics Teacher, 101*(6), 456 – 461.

Nash, T., Knowles, B., van de Wall, N., Pentz, W., Tayor, D. & Palmer, A. (2011). *Run the World (Girls)* [Recorded by B. Knowles]. On 4 [Digital download]. New York, NY: MSR Studios and Columbia.

Nuthall, G. (2004). Relating classroom teaching to student learning: A critical analysis of why research has failed to bridge the theory-practice gap. *Harvard Educational Review, 74* (3), 275.

Pellizzoni, L. (2011). Governing through disorder: Neoliberal environmental governance and social theory. *Global Environmental Change, 21*(3), 795-803.

Pinto, L.E. (2007). Textbook publishing, textbooks, and democracy: A case study. *Journal of Thought, 40*(3), 99-121.

Pinto, L.E. (2011). The F-word: How bright-sidedness overshadows feminist talk in schools. *Our Schools/Our Selves, 20*(5), 47-58.

Pinto, L.E. (2012). *Curriculum reform in Ontario: 'Common sense' processes and democratic possibilities*. Toronto: University of Toronto Press.

Pinto, L.E., McDonough, G. & Boyd, D. (2011). High school philosophy teachers' use of textbooks: Critical thinking or teaching to the text? *Journal of Curriculum and Instruction, 5* (2), 45-78.

Skovsmose, O. (1994). *Towards a philosophy of critical mathematics education*. The Netherlands: Kluwer, Dordrecht.

Smeyers, P. & Hogan, P. (2005). Inherent risks of human learning. *Educational Theory, 35*(2), 115-121.

Reynolds, C. (2012, October 31). Why are schools brainwashing our children? Protesting oil pipelines, celebrating polygamy: is the new 'social justice' agenda in class pushing politics at the expense of learning? *McLeans*. Retrieved from: http://www.macleans.ca/news/canada/why-are-schools-brainwashing-our-children/

Roberts, P. (1998). The politics of curriculum reform in New Zealand. *Curriculum Studies, 6*(1), 29-46.

Watson, A. (2006). *Raising Achievement in Secondary Mathematics.* New York, NY: Open University Press.

Wright, A. (2012). Fantasies of empowerment: mapping neoliberal discourse in the coalition government's schools policy. *Journal of Education Policy, 27*(3), 279-294.

Risky Research Business: Mathematics Education Research on the Margins

Erika C. Bullock
University of Memphis, USA

Abstract: Although we would like to believe that decisions about what research to conduct and how to conduct it are based solely on researcher interest and societal need, the reality is that external political and disciplinary factors do play a role. Scientifically based research (SBR) is one example of external political pressures that shape researcher choice both directly and indirectly. Additionally, disciplines like mathematics education operate under hidden curricula that have the potential to marginalize particular research foci. The purpose of this paper is to consider the implications of such a narrow focus on a young mathematics education researcher's choices about the lines of inquiry that she or he pursues and the ways that she or he manages the risk associated with asking and responding to questions in ways that align with SBR and other disciplinary priorities.

Keywords: mathematics education research, scientifically based research, qualitative research, research methodology, theory in research, early career research.

Scientifically based research (SBR) has been institutionalized in the United States through the United States Department of Education and the Institute for Education Sciences (IES) as the only valid or valued approach to—or the gold standard for (Lather, 2010; Schoenfeld, 2008)—education research (Popkewitz, 2006). SBR—mentioned more than 110 times in the No Child Left Behind Act of 2001 (Ritter, Anderson, Koedinger, & Corbett, 2007) and explicated by the National Research Council (2002)—is a conservative approach to education research in which social science follows a more positivist natural science model based on large-scale, double-blind, randomized, controlled trials (Lather, 2010). Eisenhart and DeHaan (2005) summarize the National Research Council's guidelines for SBR:

1. To pose significant questions that can be investigated empirically;
2. To link research to relevant theory;
3. To use methods that permit direct investigation of the question;
4. To provide an explicit and coherent chain of reasoning;
5. To replicate and generalize across studies; and
6. To make research public to encourage professional scrutiny and critique. (p. 3)

The stringent criteria used to evaluate SBR define "relevant [educational] research" as "only studies involving experimental design" (English, 2008, p. 5). These criteria dictate what programs and practices can be used in education (Shealey, 2006) and require that they be catalogued in the What Works Clearinghouse (WWC; Schoenfeld, 2008). The WWC operates, then, as a repository for interventions and strategies vetted through SBR. These strategies bear the stamp of *evidence-based*, which implies adherence to the principles of SBR. Their inclusion in the WWC positions the strategies as proven solutions to educational problems that *work* as the WWC name affirms. Thus, SBR operates under the guiding principle that the purpose of education research is to set forth solutions to education's problems (whether real or perceived) that have passed political muster, thus supporting a political agenda in education (Biesta, 2007).

The loom of SBR does not imply a dictatorial regime in which education researchers must restrict their work to prescribed methods. In fact, under SBR education researchers can pursue research questions and hypotheses in any manner they choose. However, the pursuit does entail a measure of risk as standardizing SBR necessarily marginalizes forms of research that fall outside its bounds and limits opportunities for those who do work in those forms. Mechanisms like the WWC represent the establishment of federal priorities related to the type of educational research it will support, which, by extension, dictates what types of research funding agencies will support (English, 2008). With research institutions relying more on external funding as a significant portion of tenure and promotion criteria, federal funding agencies' preferences for SBR are of particular concern to junior faculty. Ironically, as NCLB has narrowed standards for quality in education research to focus on SBR, pockets of education researchers have begun, almost simultaneously, to take up theories and methodologies that ask new questions and seek to examine old questions in new ways (Bullock, 2012; English, 2008; Hiebert, 1999). Most often, these approaches do not align with SBR's focus on uncomplicated—"too many 'on-the-other-hands' and too much attention to contextual variation and alternative possibilities produce immobilization" (Donmoyer, 2012, p. 802)—and generalizable answers to educational problems.

SBR has met little resistance in the mathematics education research community perhaps because SBR reflects the historically dominant paradigm of mathematics education research (Stinson & Bullock, 2012), a rather young and narrow research domain that has struggled to establish its identity (Kilpatrick, 1992; Schoenfeld, 2008). Although the mathematics education research community has embraced some more conservative forms of qualitative research and some use of critical social theories, the cannon of mathematics education research still aligns with a more post-positivist model. SBR's pattern of establishing strict criteria for judging what is good or valid in education research resonates within mathematics education research. Recently, Heid (2010) positioned the *Journal for Research in Mathematics Education* (*JRME*) as a space where mathematics is an "essential component rather than being a backdrop for another area of inquiry" (p. 103). While she did not explicitly disparage any particular mode(s) of inquiry in the brief editorial, the studies and areas of inquiry that she highlighted as exemplars painted a picture that reifies the historically myopic focus of mathematics education research.

Like SBR, such a statement from the then-editor of the flagship mathematics education research journal in the United States has broad implications within the field. At the time, I was a doctoral student preparing to conceptualize my dissertation research—a Foucauldian historical analysis of the National Council of Teachers of Mathematics' (1989; 2000) *Standards* documents (Bullock, 2013). I knew that I wanted to pursue an academic career and was aware that I would face certain challenges as a Black woman academic (Rockquemore & Laszloffy, 2008); I felt prepared for those battles in many ways. However, Heid's (2010) editorial comments revealed for me a new set of challenges about which I was not as confident. My research agenda did not, and still does not, align with the priorities Heid outlined for the field. "Do I have to change my interests in order to be successful?" I wondered. "Is it possible for me to achieve my goals as a mathematics education researcher as I am?" These questions continue to haunt me in waves of different concentrations. I am grateful for mathematics education scholars like de Freitas and Nolan (2008) who have provided spaces for mathematics education researchers "to take risks in their writing, to critique and disrupt cherished notions embedded in the field", thus "[troubling] many taken-for-granted assumptions about mathematics education and research" (Nolan & de Freitas, 2008, p. 1).

Early career mathematics education researchers, like me, must navigate this context of disciplinary privilege assigned to certain forms of research in order to be successful and to achieve success in the professorate as measured through high-impact publication and grant funding. We must balance a need to pursue research questions in ways that are considered valuable and fundable with a desire to do research that addresses the issues that are important to us in ways that do those issues justice and honor the communities we represent. Depending on the location of our research passions, this balance may be more difficult to locate for some researchers than others. The purpose of this paper is to consider the implications of such a narrow focus on a young mathematics education researcher's choices about the lines of inquiry that she or he pursues and the ways that she or he manages the risk associated with asking and responding to questions in ways that align with SBR.

Marginality as Practice in Mathematics Education Research

Marginalization is a process of assigning "outsider status" to anyone who "[deviates] greatly from the norms of the social organization the person is participating in or is observing," thus creating "inevitable socio emotional, political, cultural, and economic tensions" (Stanfield, 2011, p. 272). In mathematics education, we most often discuss marginalization in terms of students who experience some form of miseducation (Woodson, 1933/2008) based on race, class, gender, socioeconomic status, sexual orientation, or other factors. However, mathematics education scholars experience similar—although not nearly as grave—effects when facing external and internal structures that attempt to direct mathematics education research by establishing norms shrouded by "the myth of neutrality" (Martin, Gholson, & Leonard, 2010, p. 13).

Constructing Margins from Without

The increased acceptance of qualitative methodologies in mathematics education research combined with mathematics' firm position at the center of education policy conversations for several decades (Steiner, 1987) have brought external challenges to the "quality and usefulness of mathematics education research" (Simon, 2004, p. 157). Legislation such as NCLB and stipulations like SBR place constraints on the types of research that education researchers conduct. As the government dictates to education researchers what qualifies as science, these constraints force scholars to make a decision about whether they will (a) move in lockstep with governmental priorities; (b) resist the pressure to embrace SBR and continue to do research as they see fit; or (c) find ways to negotiate SBR's constraints.

Federal funding agencies such as the United States Department of Education and the National Science Foundation align their funding priorities with legislative priorities. For example, these agencies have established initiatives to support colleges of education in training doctoral students explicitly in SBR (Eisenhart & DeHaan, 2008), thus influencing how researchers are trained. In addition, the funding environment influences how researchers do their work (Zoller, Zimmerling, & Boutellier, 2014). In an era of increased pressure for university faculty to secure external funding as part of tenure and promotion requirements, the priorities that these agencies establish can serve as drivers for the ideas that junior faculty pursue in order to secure a tenured position in the academy. A junior scholar may ask of her or his original idea "Is this idea fundable?" and, if the answer is no, choose to set the idea aside for later investigation, possibly post-tenure. Alternatively, she or he may use funding guidelines and priorities as the framework within which she or he asks questions by asking, "What ideas are fundable?" and choosing research projects accordingly. Although neither of these scenarios is dire, the potential for delayed pursuit or abandonment of questions that are important to the researcher and to the field is a loss for both.

Constructing Margins from Within

Although external factors such as federal education policy contribute significantly to the marginalization of research outside of SBR's purview, marginalization within mathematics education has not been the sole work of external forces. In many ways, mathematics education has participated in its own exclusionary practices. In recent years, prominent mathematics education scholars have addressed some of the discipline's exclusions. The three scenarios that follow represent current and persistent tales of marginalization occurring within the community of scholars interested in mathematics teaching and learning.

First, Parks and Schmeichel (2012) investigated the limitations of *JRME* in addressing issues of equity—and particularly race and ethnicity—as a central issue within mathematics education. The motivator for the commentary was a review of literature in *JRME* that revealed minimal efforts to address race and ethnicity in substantive ways. They conclude:

> The marginalization and exclusion of attention to race and ethnicity in mathematics education discourse is problematic because what we write both reflects and shapes what we know and believe to be true about the field. The absence of engagement with ideas of power, identity, and equity in mathematics education research reiterates a regulatory schema (Butler, 1993) that inhibits thinking about these forces as relevant to learning in the mathematics classroom and, by extension, limits our capacity for thinking about power and identity in ways that can make a different for students. (p. 250)

These observations demonstrate a lack of willingness to engage the ways in which power and privilege operate within mathematics education.

As a second example, Heid's (2010) *JRME* editorial "Where's the math (in mathematics education research)?" mentioned earlier prompted a response from Martin, Gholson, and Leonard (2010)—and subsequent response from Battista (2010) and Confrey (2010)—that pointed out some of mathematics education's exclusionary practices and its reluctance to acknowledge the inherently political nature of *all* inquiry in mathematics education. Heid (2010) argued that research in mathematics education should "focus on critical features of mathematical understanding" and that "*JRME* publishes research in which mathematics is an essential component rather than being backdrop for another area of inquiry" (p. 103). According to Martin and colleagues (2010):

> [Heid's] statements also represent very public displays of power and privilege. The implications for such exercises of power, under the auspices of an institutional and organizational entity such as the NCTM [National Council of Teachers of Mathematics], are profound, as they have the potential to marginalize scholarship within particular areas of focus as well as to marginalizes scholars who devote themselves to this work. Young scholars and graduate students are particularly vulnerable if the subtext of these statements is on pursuing what is valued in the field, as decided by those in positions of power, versus choosing what they want to make their life's work. (pp. 13−14)

Battista (2010) asserts that inherent to Heid's and Martin and colleagues' arguments could be the question "What is mathematics education research?" (p. 37) but even this question is insufficient as it ignores the research in policy, cognitive psychology, and other areas that allow a better understanding of mathematics education and its elements. Thus, he proposes a more appropriate question: "What kinds of research is [*sic*] needed for mathematics educators to understand and improve mathematics learning and teaching?" (p. 38). This question opens up more possibilities for mathematics education research, but it does not seem that the field has taken up these questions in earnest.

Finally, Boaler's (2012) discussion of marginalization reveals direct attacks from mathematicians who attempted to discredit her work. She outlines the disagreements that two mathematicians registered with her work and a history of moves on their behalf to discredit that work. Boaler garnered support from the mathematics education community around the globe, as evidenced by the more than 1,000 signatures on a petition demanding action from Stanford University and decrying the mathematicians for undermining the integrity of the academy.[1] Although Boaler's case represents a reprehensible act of academic bullying (Stinson, 2012), it also brings to light lingering tensions—or math wars—between the mathematics and mathematics education communities about who has the authority to make recommendations about school mathematics (see e.g. Jackson, 1997; Schoenfeld, 2004; Stinson, 2012).

Opportunity Amidst Chaos

Schoenfeld (2008) described mathematics education as a young discipline that is going through a chaotic period as it works to locate its identity. There are two options for mathematics education researchers to respond to this chaos. We can perceive the chaos as a negative consequence of an unfortunate growth period and hope that the field quickly settles into a sense of "normal." The alternative is to view the chaos positively as an opportunity to create mathematics education in various forms in response to the issues that face the discipline and to hope that chaos becomes accepted as "normal" in the discipline. I choose the latter; I see the chaos and instability as a window of opportunity, albeit narrow, to think of the "network of mathematics education practices" (Valero, 2012, p. 374) in new and different ways. The chaos excites me and propels me forward as a scholar in mathematics education, but that excitement is tempered by an ever-present reality that the questions I want to ask and the theories and methodologies I want to use in response are not necessarily commensurable with the discipline's current priorities.

It has been written that in this current state of chaos, mathematics education research is driven by the researcher's whims (Lesh & Lovitts, 2000). I would prefer to assign the motivation for inquiry to researcher experiences and curiosities. Like many researchers, my research questions come from my experience as a mathematics student, teacher, and education researcher. As a secondary mathematics teacher in schools where students were labeled as *at risk* and isolated from high-quality mathematics experiences, I often wondered how mathematics education came to the point where we fought rhetorically for *mathematics for all* (Martin, 2003), yet classrooms like mine were filled with Black students that were locked out. These questions and experiences shape my research agenda. My dissertation research (Bullock, 2013), for example, began with a question of how one organization was able to provide recommendations for school mathematics (i.e., the NCTM [1989; 2000] *Standards*) that shaped approaches to mathematics throughout the United States for two decades or more. This work provides a historical backdrop and has generated new questions. Conceptualizing urban mathematics education—my current research focus—has become a passion that drives me toward urban sociology, critical geography, urban planning, and other disciplines to begin to better understand how the urban space shapes students' mathematics experiences. I know as a teacher and teacher educator that there are nuances to urban education and believe that it is important to engage these nuances within mathematics education (Howard & Milner, 2013; Martin & Larnell, 2013; Tate, 2008). My research is neither traditional nor aligned with SBR, but it is a contribution. Therefore, describing it as a result of whims or unexplained desires devalues these experiences and encounters. It reduces non-traditional approaches to mathematics education research to the object of researcher self-interest rather than efforts to advance the discipline through substantive inquiry.

Unfortunately, mathematics education's position of importance related to education policy in the United States (Steiner, 1987) subjects mathematics education research to political pressure. Therefore, issues important to federal agencies rather than those important to researchers drive inquiry in

---

[1] Jeremy Kilpatrick, University of Georgia started the petition: http://www.change.org/p/the-community-of-mathematics-educators-join-in-defending-fundamental-values?utm_campaign=action_box&utm_medium=twitter&utm_source=share_petition

mathematics. When we reduce novel research approaches to researcher whims, attempts to bring attention to the dearth of alternative literatures are likened to the complaints of an irritated child rather than a call for open and inclusive scholarly discourse.

Early career researchers in mathematics education must measure the internal and external marginalization of ideas with institutional demands. Increasingly, universities expect junior faculty to secure external funds as a prerequisite for tenure, particularly in fields like mathematics education in which thee funds flow more freely than others. Additionally, tenure-seeking faculty are expected to publish research in the leading journals of the field. Facing the restrictions of SBR on funding priorities and the internal pressure to conduct research that can be considered more internally valid, what is the early career mathematics education researcher to do when she or he finds her or his work to be in the margins? As mentioned early, there are three options: (a) follow along with governmental priorities; (b) resist the pressure to embrace favored approaches; or (c) negotiate the constraints. The first option can constitute a denial of the researcher's interests and passions. The second option can cause the researcher to experience unfavorable tenure and promotion decisions. The final option, negotiating the demands and preferences for traditional research with the researcher's desire to ask new questions, seems to be the most reasonable response. This process of negotiation requires the mathematics education researcher to frame her or his research in ways that are more palatable to external funders. Her or his publication strategy must include interdisciplinary journals that embrace the work without avoiding mathematics education venues. The only way that publication venues in mathematics education can expand their boundaries is if scholars continue to submit good work. Eventually, the opportunity will come. Choosing to play in the proverbial sandbox of mathematics education instead of relegating oneself to the margins is not easy. It requires patience, persistence, high tolerance for rejection, and a willingness to redefine success. However, I believe that the greatest motivator is love.

With Love for Mathematics Education

In a seminar with early career faculty mathematics education during the Association of Mathematics Teacher Educators' Service, Teaching, and Research (STaR) mentoring institute, Sandra Crespo (2014) asked participants to consider how love motivates our research. After some thought I realized that my love of mathematics and mathematics education motivates my scholarly approach; I love mathematics education as a discipline too much to allow it become stagnant in any way. I love mathematics and mathematics education too much to allow it to be deprived of the creative energy that those who have been historically excluded from full access can bring. I love mathematics and mathematics education enough to pursue any and all intellectual means to contribute to its expansion and improvement. I love mathematics and mathematics education enough to push in an effort to shore up its weaknesses and to reveal its potential.

But love is not without risk. My love for mathematics and mathematics education prompts me to ask questions and to use methodologies that may not align with the canon of the discipline and may position me in the margins. These questions and methodologies do not always fall on the right side of the "Where's the math?" (Heid, 2010) questions asked within our most prominent publication venues or engage the types of research design favored by funding agencies that prioritize SBR. Unfortunately, love of mathematics and mathematics education is not enough to guard against calls to justify the scientific value of investigating the power relations at work in curriculum and policy making or proposing different theoretical and methodological approaches for mathematics education research. Love of children is not enough to address the persistent inequities that mar their mathematics experiences *by any means necessary.* But perhaps it should be.

References

Battista, M. T. (2010). Engaging students in meaningful mathematics learning: Different perspectives, complementary goals. *Journal of Urban Mathematics Education*, *3*(2), 34–46.

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *51*(1), 1–22.

Boaler, J. (2012, October). Jo Boaler reveals attacks by Milgram and Bishop: When academic disagreement becomes harassment and persecution. Retrieved from http://web.stanford.edu/~joboaler/

Bullock, E. C. (2012). Conducting "good" equity research in mathematics education: A question of methodology. *Journal of Mathematics Education at Teachers College*, *3*(2), 30–36.

Bullock, E. C. (2013). *An archaeological/genealogical historical analysis of the National Council of Teachers of Mathematics standards documents* (Doctoral dissertation). Georgia State University, Atlanta, GA. Retrieved from http://scholarworks.gsu.edu/msit_diss/110/

Confrey, J. (2010). "Both And"—Equity *and* mathematics: A response to Martin, Gholson, and Leonard. *Journal of Urban Mathematics Education*, *3*(2), 25–33.

Crespo, S. (2014). Researching with love and care: The journey to becoming an intentional and reflective researcher. Presented at the Service, Teaching, and Research Institute, Park City, UT.

de Freitas, E., & Nolan, K. (Eds.). (2008). *Opening the research text: Critical insights and in(ter)ventions into mathematics education*. New York, NY: Springer.

Donmoyer, R. (2012). Two (very) different worlds: The cultures of policymaking and qualitative research. *Qualitative Inquiry*, *18*(9), 798–807.

Eisenhart, M., & DeHaan, R. L. (2005). Doctoral preparation of scientifically based education researchers. *Educational Researcher*, *34*(4), 3–13.

English, L. D. (2008). Setting an agenda for international research in mathematics education. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 3–19). New York, NY: Routledge.

Heid, M. K. (2010). Where's the math (in mathematics education research)? *Journal for Research in Mathematics Education*, *41*(2), 102–103.

Hiebert, J. (1999). Relationships between research and the NCTM Standards. *Journal for Research in Mathematics Education*, *30*(1), 3–19.

Howard, T. C., & Milner, H. R. (2013). Teacher preparation for urban schools. In H. R. Milner IV & K. Lomotey (Eds.), *Handbook of urban education* (pp. 199–216). New York, NY: Routledge.

Jackson, A. (1997). The math wars: California battles it out over mathematics education reform (part I). *Notices of the American Mathematical Society*, *44*(6), 695–702.

Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 3–38). Reston, VA: Macmillan.

Lather, P. (2010). *Engaging science policy: From the side of the messy*. New York, NY: Peter Lang.

Lesh, R., & Lovitts, B. (2000). Research agendas: Identifying priority problems, and developing useful theoretical perspectives. In A. E. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 45–72). Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, D. B. (2003). Hidden assumptions and unaddressed questions in *mathematics for all* rhetoric. *The Mathematics Educator*, *13*(2), 7–21.

Martin, D. B., Gholson, M. L., & Leonard, J. (2010). Mathematics as gatekeeper: Power and privilege in the production of knowledge. *Journal of Urban Mathematics Education*, *3*(2), 12–24.

Martin, D. B., & Larnell, G. V. (2013). Urban mathematics education. In H. R. Milner IV & K. Lomotey

(Eds.), *Handbook of urban education* (pp. 373–393). New York, NY: Routledge.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academies Press.

Nolan, K., & de Freitas, E. (2008). Foreword to the research text: Mathematics education under cross-examination. In E. de Freitas & K. Nolan (Eds.), *Opening the research text: Critical insights and in(ter)ventions into mathematics education* (pp. 1–11). New York, NY: Springer.

Parks, A. N., & Schmeichel, M. (2012). Obstacles to addressing race and ethnicity in the mathematics education literature. *Journal for Research in Mathematics Education*, *43*(3), 238–252.

Popkewitz, T. S. (2006). Hopes of progress and fears of the dangerous: Research, Cultural theses, and planning different human kinds. In G. J. Ladson-Billings & W. F. Tate IV (Eds.), *Educational Research in the Public Interest* (pp. 119–140). New York, NY: Teachers College Press.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin Review*, *14*(2), 249–255.

Rockquemore, K. A., & Laszloffy, T. (2008). *The Black academic's guide to winning tenure—without losing your soul*. Boulder, CO: Lynne Rienner Publishers, Inc.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, *18*(1), 253–286.

Schoenfeld, A. H. (2008). Research methods in (mathematics) education. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 467–519). New York, NY: Routledge.

Shealey, M. W. (2006). The promises and perils of "scientifically based" research for urban schools. *Urban Education*, *41*(1), 5–19.

Simon, M. A. (2004). Raising issues of quality in mathematics education research. *Journal for Research in Mathematics Education*, *35*(3), 157–163.

Stanfield, J. H., II. (2011). *Historical foundations of black reflective sociology*. Walnut Creek, CA: Left Coast Press.

Steiner, H.-G. (1987). Philosophical and epistemological aspects of mathematics and their interaction with theory and practice in mathematics education. *For the Learning of Mathematics*, *7*(1), 7–13.

Stinson, D. W. (2012). Mathematics educators and the "Math Wars": Who controls the discourse? *Journal of Urban Mathematics Education*, *5*(2), 1–5.

Stinson, D. W., & Bullock, E. C. (2012). Critical postmodern theory in mathematics education research: A praxis of uncertainty. *Educational Studies in Mathematics*, *80*(1–2), 41–55.

Tate, W. F. (2008). Putting the "urban" in mathematics education scholarship. *Journal of Urban Mathematics Education*, *1*(1), 5–9.

Valero, P. (2012). A socio-political look at equity in the school organization of mathematics education. In H. Forgasz & F. Rivera (Eds.), *Towards Equity in Mathematics Education* (pp. 373–387). Berlin, Germany: Springer.

Woodson, C. G. (2008). *The mis-education of the Negro*. New York, NY: Classic House Books. Original work published in 1933.

Zoller, F. A., Zimmerling, E., & Boutellier, R. (2014). Assessing the impact of the funding environment on researchers' risk aversion: the use of citation statistics. *Higher Education*, *68*(3), 333–345.

Risk as an Explanatory Factor for Researchers' Inferential Interpretations

Rink Hoekstra
University of Groningen, Netherlands

Abstract: Logical reasoning is crucial in science, but we know that this is not something that humans are innately good at. It becomes even harder to reason logically about data when there is uncertainty, because there is always a chance of being wrong. Dealing with uncertainty is inevitable, for example, in situations in which the evaluation of sample outcomes with respect to some population is required. Inferential statistics is a structured way of reasoning rationally about such data. One could therefore expect that using well-known statistical techniques protects its users against misinterpretations regarding uncertainty. Unfortunately, this does not seem to be the case. Researchers often pretend to be too certain about the presence or absence of an effect, and data are analysed in a selective way, which impacts the validity of conclusions that can be drawn from the techniques that are used. In this paper, the concept of risk is used to explain why unwanted behaviour may not be as unreasonable as it seems, once the risks that researchers face are taken into account.

Keywords: risk, inference, statistics, questionable research practices, uncertainty.

Prelude

Risk may not be the first thing that comes to mind when trying to explain how researchers deal with uncertainty in their papers when talking about their data. Nevertheless, I think it may play a larger role than one might expect. For this paper, I will use Kaplan and Garrick's (1981) quantitative interpretation of risk, in which they present risk analyses as an answer to a set of three questions:

(i) What can happen?

(ii) How likely is it that this will happen?

(iii) If it does happen, what are the consequences?

In explaining how researchers deal with uncertainty, the concept of risk is relevant from two different angles. First, interpreting inferential statistical outcomes necessarily involves uncertainty. Every now and then, one happens to end up with a relatively atypical and therefore not representative sample outcome, and the consequence could be that future studies on the same issue will find something quite different from what you have found, which could cause the earlier conclusions to be called into question. In this paper, however, the focus will be mainly on risk from a second angle. From this angle, the focus is on the personal risks that might impact publication of a scientific paper and the reputation of a researcher when making certain choices regarding the reporting and interpretation of the outcomes. Once the likeliness of undesirable consequences resulting from these choices is taken into account, seemingly unreasonable behavior like presenting data too certainly, becomes more understandable. In this paper, I will argue that understanding risk aversion may be an important step for being able to improve what some refer to as a crisis in the social sciences.

Reasoning with uncertainty

It is well known that human reasoning is far from optimal, and it becomes even harder when uncertainty is involved. Tversky and Kahneman (1974) showed that when dealing with uncertainty, we often rely on a relatively small number of heuristics to come up with an answer. Gigerenzer (2008) argues that it is important to take the context in which conclusions are drawn into account, to understand why interpretational mistakes that seem illogical at first sight, may make sense once the context is understood.

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 103-112

Some situations that involve a human interpretation under uncertainty require an answer as rational as possible given the data at hand. For example, one would expect a counsellor with expertise regarding a certain illness to be able to interpret results of tests for that illness. Gigerenzer, Hoffrage and Ebert (1998) showed for a sample of AIDS counsellors in Germany that this was not the case. False positives (testing positive on a test when you do not have the disease that is tested for) were often not taken into account, and positive results were sometimes presented as conclusive evidence that the one who was tested had the disease. We also expect judges to make sound judgements about the likeliness or unlikeliness of the accused being guilty. Unfortunately, this is also not always true. In a famous English case from 1999, Sally Clark was convicted for murder of two of her sons, who had been found dead in their beds. The sudden infant death syndrome (a residual category in case all other explanations have been excluded) was ruled out as highly unlikely, but the unlikeliness of a mother killing two of her sons was not taken into account (Manktelow, 2012). In 2003, a Dutch nurse called Lucia de Berk was convicted on the basis of only circumstantial evidence for the murder of four patients and attempted murder of three others, because it was ruled to be too coincidental that she was on duty when the patients died (Buchanan, 2007). Again, the unlikeliness of a nurse killing or trying to kill several patients was not taken into consideration. Eventually, de Berk was imprisoned for 5 years, and after being released in 2008 pending a higher court decision, the court ruled in 2010 that the defendant was not guilty. The examples above show that in some cases, the misinterpretation of uncertainty can have an enormous impact.

As from counsellors and judges, we expect scientific researchers to draw justified conclusions based on the data at hand. In a scientific context, a researcher who tries to answer a certain research question often has a limited amount of data to make valid inferences. When researchers want to know something about a certain group (the population), they are often restricted to information of only a small number of people from this population (the sample). This implies that the eventual outcomes are only based on a small subset of the entire group in which the researcher is interested. In these cases, uncertainty is intrinsically connected to any statement that is made about the population. For example, if a researcher wants to study which proportion of the population that is entitled to vote would vote for candidate C if the election would be held today, it is practically impossible to ask everyone in the population. Instead, one would aim at selecting a random sample of participants from the population, and try to estimate the value in the population one is interested in (which is called the parameter). Assuming a completely random sample, and assuming completely cooperative and honest participants, the best possible estimate based on the sample would be the proportion of people *within* the sample that would indicate voting for C. But it is still unsure what the exact value for the parameter is: Despite random selection, the sample might not show *exactly* the same proportion as the population. Intuitively, if the sample were relatively large, one would expect the sample outcome to be close to the parameter, but it is unclear how close. How can we make a reasonable interpretation of the parameter based on the outcome, knowing how hard it is for people to reason with uncertainty? To overcome the human deficits with respect to reasoning about data with uncertainty, a systematic approach of analysing data seems necessary.

Inferential statistics are designed to provide a structure to reason rationally about sample data, in order to make an inference about the parameter. Given that statistics is a means to deal with uncertainty in a structured way, one could expect that when researchers use statistics for their inferential statements, reasoning with uncertainty is done without too many problems. The question is whether this is really the case in practice.

Two inferential statistical frameworks

For reasons of clarity, it is necessary to present the two most commonly used inferential statistical frameworks in the social sciences: Frequentist statistics (which is sometimes also referred to as classical statistics, or frequentism) and Bayesian statistics. The former is by far the most frequently used framework, whereas the latter seems to gain momentum in the last couple of decades, and some even refer to it as a Bayesian revolution (e.g., Brooks, 2003; Zyphur & Oswald, 2015).

Frequentism defines probability as a long-term relative frequency (Dienes, 2008). The two frequentist techniques that are most often used are the null hypothesis significance test (hereafter referred to as significance test) and the confidence interval (CI). In the social sciences, biology, and the medical sciences, significance tests can be found in a large majority of the published papers. A significance test, as it is most often used, calculates the probability of finding the outcome found in the sample or more extreme, under the assumption that a specific test-value you are interested in equals a certain value (often null, hence null-hypothesis significance testing). If this probability, the *p*-value, is smaller than a certain criterion value called the significance level (which is often chosen to equal 5%), the outcome is called "significant", and if it is larger, the outcome is considered "not significant". The significance level indicates the risk of incorrectly rejecting the null hypothesis (also referred to as Type I error). Note that "significant" in this context is only a statistical term: It does not refer to the practical importance of the finding, but it merely indicates how likely the outcome or more extreme is given a certain value of the parameter. In order to say something about practical relevance, power calculations could indicate how "powerful" (strong) the procedure at hand is. This depends on, amongst others, the sample size, and the homogeneity of the group at hand. The power indicates how likely it is to find a significant effect under the assumption that in reality a certain value for the parameter were true. The complement of this probability is called the Type II error.

A second technique within the frequentist framework is the confidence interval. The idea behind CIs is that you construct a certain interval around the value you found in your sample in such a way, that, if an infinite number of independent samples were taken from the same population and a confidence interval was constructed for each sample, a certain percentage (for example 95%) of these intervals would include the parameter. CIs are often endorsed as an alternative for the significance test, since they are assumed to give an indication of precision of the parameter estimate, and since they give a direct indication of the size of the effect (Cumming, 2012).

To complicate matters even further, the standard way of using these techniques in practice deviates from theory, and there are at least two theories about which techniques should be used, and how they should be interpreted. Historically, Fisher (e.g., Fisher, 1925) considered drawing conclusions as an essential part of inference, and according to him *p*-values are to be considered as measures of evidence. In his model, no alternative hypotheses are considered. Neyman and Pearson extended his model, but they argued that frequentist inference is to be used for decisions, and not for drawing conclusions. They also stressed the importance of keeping long-term error rates low, thus focusing on the importance of the significance level and power. The debates between these two camps were far from friendly. Ironically, the current model that is typically used seems a hybrid version of Fisher's and Neyman and Pearson's view on frequentist statistics (Gigerenzer, 1993), and it would probably have been rejected by both sides, although on different grounds (Kline, 2013). Because of its prevalence, I will focus on the hybrid version that was introduced before. The reader should keep in mind, however, that depending on the position one has, the interpretations that are considered correct and not correct may differ.

Probability is defined differently within the Bayesian framework. The core idea of Bayesian statistics is that it quantifies how a rational person would change his or her prior beliefs about a parameter based on the data at hand. That is, contrary to frequentist statistics, Bayesian statistics require a specification of one's belief about plausible values of the parameter before looking at the data. Within Bayes, there are several ways of analysing data. First of all, you can update the prior beliefs by means of the data. That is, given a quantification of your prior idea of the position of the parameter, Bayesian statistics tell you how you should update your belief based on the data if you were a completely rational person. An alternative way is to contrast two models of the data, and compare the likelihood of both models given the data. This ratio of likelihoods is called the Bayes Factor (Dienes, 2011).

Now that the two main inferential frameworks (frequentism and Bayesianism) have been introduced, the question is whether they are usable for researchers to draw sound conclusions. How do researchers in practice deal with uncertainty when making inferential claims? What do we know about their use of statistical methods? Do the statistical techniques they use protect them from making reasoning errors? Or do they evoke reasoning errors themselves? In the sequel, it will be discussed how people use inferential statistics to reason about data. In a second step, it will be explored how the definition of "risk" as it was introduced earlier, may play an important role in explaining why people use statistics as they do.

The use of inferential techniques in practice

Analysing and interpreting data by means of inferential techniques involves many aspects. For the purpose of this paper, we focus on the following two questions:

1) How do researchers acknowledge the uncertainty that is inextricably connected to inference in their conclusions?

2) Does the way people analyse the data justify the use of the technique at hand?

For answering the first question, I will focus on our knowledge of how people report and interpret their statistical outcomes. For the second question, I will show that given how many researchers analyse their outcomes selectively, the conclusions based on the techniques are often invalid. Since a majority of the studies that are presented had researchers in the social sciences as participants, the scope of the eventual conclusions is limited to the social sciences as well.

Numerous articles have been written about the usability (or the lack thereof) of frequentist and Bayesian techniques (for an overview, see for example Kline, 2013). The debates between frequentists and Bayesians have been (and are to this very day) typically rather fierce, but also within both frameworks quite some disagreement can be found. Despite the attention for the usability of both techniques, few studies have focused on their use in practice.

As stated before, significance testing is by far the most frequently used technique within the social sciences, medical sciences, biology and beyond (Bakker, van Dijk, & Wicherts, 2012; Fidler & Loftus, 2009; Hoekstra, Finch, Kiers, & Johnson, 2006; Hubbard & Ryan, 2000; Kline, 2013; Sterling, Rosenbaum, & Weinkam, 1995). Unfortunately, despite the earlier mentioned "Bayesian revolution", I am not aware of any study on the use in practice of Bayesian techniques. In the sequel, the focus will therefore be on the use of frequentist statistics, which should by no means be interpreted as if frequentism were the only way to analyse data. In the absence of information about the use of Bayesian methods, however, I cannot draw any conclusion about its use. In the conclusion section I will return to this by discussing to what extent Bayesian statistics might be a solution for the presented problems.

As to how uncertainty is acknowledged in interpretations (the first question), it is known that the conclusions that are drawn from NHST are often not without errors. Significant results are often interpreted as if it is indisputable that the null hypothesis is false, thus ignoring the risk of making a Type I error, and non-significant results are often presented as if they reflect the null hypothesis being true, thus ignoring the risk of making a Type II error (Hoekstra et al., 2006). In Hoekstra et al.'s paper, interpretational errors were defined leniently. That is, frequentists who adhere to the Neyman-Pearson interpretation of frequentist statistics could argue that only *decisions* but no *conclusions* are justified given frequentist outcomes. From that perspective, *any* conclusion would be unwarranted, but in this paper, only overly rigid conclusions were counted.

When possible interpretations of *p*-values are presented to researchers, very few are able to correctly indicate which conclusions are justified (Falk & Greenbaum, 1995; Oakes, 1986). Haller and Kraus (2002) showed that even statisticians endorsed incorrect statements about *p*-values. Similar findings have been found for statisticians at pharmaceutical companies (Lecoutre, Poitevineau, & Lecoutre, 2003) and doctors and dentists (Scheutz, Andersen, & Wulff, 1988; Wulff, Andersen, Brandenhoff, & Guttler, 1987).

Confidence intervals, on the other hand are seldom reported (Cumming, 2012; Fidler & Loftus, 2009; Hoekstra et al., 2006), and researchers seem to have great difficulty with interpreting them correctly as well (Hoekstra, Morey, Rouder, & Wagenmakers, 2014). When the same data are presented by means of significance testing or CIs in an experimental setting, it was found that both are interpreted differently (Hoekstra, Johnson, & Kiers, 2012). Since the presented techniques conveyed exactly the same information, it was only presented differently, logical reasoning should have led to exactly the same conclusion. Belia, Fidler, Williams, and Cumming (2005) also showed that researchers were relatively bad in showing awareness of the relation between significance testing and CIs.

The second question deals with the way data are treated before the analysis that is used in the version that is eventually submitted for publication in a journal. John, Loewenstein and Prelec (2012) showed that when analysing data, researchers often selectively adjust their analyses based on the significance of their findings. That is, in case of a significant effect they write down their conclusions without further ado, whereas non-significant findings may lead to adjustments like increasing the sample size, or leaving out one of the conditions or dependent variables, amongst others. They dubbed such adjustments questionable research practices (QRPs). Another QRP is presenting a significant outcome as if it were predicted beforehand, whereas it actually was an unexpected finding. By means of simulations, Simmons, Nelson, & Simonsohn (2011) found that using such QRPs selectively (that is, only or mainly in case of non-significant findings) inflates the Type I error rate unacceptably. As a result, the interpretation of a *p*-value or the significance of the outcome is practically impossible, and we have shown earlier that researchers find this already difficult enough without QRPs. The most extreme QRP is fabricating data, which is obviously wrong. Some of the QRPs, however, are so generally accepted that most researchers are not even aware that these practices are questionable. Note that this implies that the researcher who exhibits QRPs does not necessarily have to be aware that this is problematic. All in all, there are strong indications that QRPs are part of the routine of researchers. This has serious implications for the interpretation of the results, since the outcomes do no longer mean what they originally mean. In fact, outcomes like *p*-values and CIs are almost impossible to interpret if we would allow the use of QRPs.

In summary, frequentist inferential statistical outcomes are often interpreted incorrectly (even when using lenient standards), and in papers effects or the alleged lack of an effect are often presented with more certainty than is justifiable. Moreover, the apparent frequent use of QRPs impacts the interpretability of the outcomes even further. Apparently, despite the fact that inferential statistics have been designed to prevent interpretational errors from occurring, they are no guarantee for the absence of such mistakes in the published literature.

Risk as an explanatory factor

In this paragraph, I will try to show how the previously mentioned use and interpretation of inferential outcomes can be partially explained by risk aversion behaviour. In order to assess researchers' considerations with respect to the aforementioned choices, I will use Kaplan and Garrick's (1981) risk analysis by answering the set of three questions: *What can happen?*, *How likely* *is it that this will happen?*, and *If it does happen, what are the consequences?*. Although I am fully aware of the fact that researchers are part of a community in which behaviour of one person influences the behaviour of someone else, the focus in this paragraph will be on the choices that the individuals make within the scientific system. These choices are subsequently related to the context in which they are made.

The idea that risk aversion may explain why the unwanted behaviours (intentionally or not) like misinterpreting statistical outcomes and using QRPs occur is not a completely new insight. Many have argued that the incentive structure in academia is at the core of the problem that we are having in psychology and beyond (e.g., Asendorpf et al., 2013; Nosek, Spies, & Motyl, 2012; Nosek & Lakens, 2014; Open Science Collaboration, 2012). Avoiding risks can be considered complementary to following incentives. With respect to risks, two things are of crucial importance for being a successful researcher: Having a decent amount of publications, and having a solid scientific reputation.

Let's first focus on the risks involved with one of the types of unwanted behaviour discussed in this paper: Using one or more QRPs when the original finding, after analysing the data as was originally planned, was not significant. In the current scientific publication system (and assuming the use of frequentist statistics), a significant finding seems a requisite for a publishable paper. As a result, researchers are typically disappointed when their effects happen to be non-significant. At this point, QRPs may be tempting. When *not* using them in case of a non-significant finding (and according to Cohen (1994), non-significant findings should be found regularly, since power is typically low), the risk of *not* publishing the article is high. Nevertheless, non-significant findings could be indicative of interesting findings. For example, in case of many studies with low power, a series of low-powered studies could result in a very clear effect in a meta-analysis, in which outcomes of several studies are combined. If papers with non-significant effects are not presented, however, we will end up with a so-called *file-drawer problem* (non-significant findings disappear in a file-drawer), which makes it hard to get a good estimate of the parameter. There are published papers to be found for which findings were non-significant, but this is relatively rare. The use of QRPs does definitely increase the probability of an article getting published, since the probability of finding a significant effect is artificially increased (and significant results have a higher probability of being published). Thus, avoiding QRPs might lead to a paper not being published, and it consequently might hamper a researcher's career and reputation. The probability of that reputation being hampered in case a researcher decides to use QRPs, on the other hand, is negligible, because the probability of getting caught when doing so is practically zero: Typically, a researcher is not being observed when analysing the data. And even if someone happens to get caught, the reputational damage is limited: The argument "Everybody else is doing it, so why can't I?" can easily be used as a defence. Of course, this does not hold for the use of QRPs like fabricating data. If this comes out, the researchers' reputation will be irreparably damaged, but many other QRPs seem to be considered acceptable.

A second choice a researcher has to make considers acknowledging uncertainty. Within the context of significance testing, ignoring uncertainty results in interpreting a significant effect as definite proof that the null-hypothesis is untrue (thus ignoring the possibility of a Type I error), and a non-significant effect as if there is definite proof that the null-hypothesis is true (thus ignoring the possibility of a Type II error). A pragmatic researcher would base this choice at least partly on the impact it might have on the publishability of the paper and on his or her reputation. If uncertainty is acknowledged, the paper is somewhat more honest, but also harder to read, and arguably less convincing. It is, for example, easier to understand "A is explained by B, but not by C" than a more truthful "A may be explained by B, and we don't know yet whether C plays a role". The culture in academia seems to require novel and groundbreaking research, and questioning oneself could be seen as a sign of weakness. In that sense, acknowledging uncertainty can decrease the probability of a paper being published, and if it is published, it can be seen as if the results are not very strong, which can impact the reputation of the researcher negatively, although the latter risk is most likely smaller than the former. Presenting the results in a way that implies more certainty than is justified, on the other hand, does not seem risky at all. The likeliness of being accused of being too certain is slim, since it is common practice among behavioural scientists, and since the papers become slightly more difficult the read when uncertainty is acknowledged. If someone would point it out though, the implications for the reputation are most likely negligible.

In summary, we have seen that mistreating data (QRPs) and a misinterpretation of outcomes (ignoring uncertainty) are observed regularly. Although unwanted, this behaviour can be considered rational once the scientific context in which the researcher works including the risks within that context is taken into account. The risks of not publishing an article or of a damaged reputation seem higher when QRPs are *not* used, and, arguably, the risks are also higher when uncertainty is acknowledged.

Conclusion and Discussion

Researchers have difficulties to reason soundly about their inferential outcomes. Although we know that people are generally bad in reasoning with uncertainty (see e.g., Gigerenzer, 2008; Tversky & Kahneman, 1974), one would hope that inferential techniques, which are designed to assist people in reasoning in a structured way, would prevent these problems, but clearly that is not the case. Not only do researchers often seem to ignore the amount of uncertainty that is inextricably connected to every conclusion that is drawn about a population based on a sample outcome, but the regular use of QRPs makes it almost impossible to draw valid conclusions from the frequentist outcomes that are typically used. QRPs and pretending certainty are two different issues, but they are connected: With QRPs the probability of finding a Type I error are artificially inflated, and when uncertainty is not acknowledged Type I and Type II errors are basically ignored.

The findings that QRPs are regularly used and that uncertainty is often not acknowledged are non-trivial outcomes: In some papers the current status of the use of inferential methods in psychology is referred to as a “crisis” (e.g., Pashler & Wagenmakers, 2012), and replication studies show that many significant outcomes are not significant in exact replications of the studies (e.g., Galak, LeBoeuf, Nelson, & Simmons, 2012; Harris, Coburn, Rohrer, & Pashler, 2013), which could be indicative for a regular use of QRPs in the original articles. Unfortunately, knowing that QRPs are sometimes used also impacts the trustworthiness of those papers for which they were not used, since it is almost impossible for a reader to distinguish between the two.

As stated before, the explanation for this undesirable behaviour can most likely not be reduced to one factor only. Time constraints, lack of knowledge of the finesses of the statistical techniques that are used, the inclination to adjust to the behaviour of others and the power of habit, amongst others, all play a role in this process. In this paper, however, the focus is on the role of risk in the choices that individual researchers seem to make. From the perspective of risk aversion, the unwanted behaviour may not be as illogical as it seems. If one wants to minimize both the risk of a hampered reputation and the risk of not publishing a sufficient amount of papers, presenting outcomes as clear as possible, -even if this impacts the formally correct interpretation somewhat-, and using some QRPs in case the key outcomes happen to be non-significant seem in fact rational choices.

If we, as a scientific community, consider such behaviour unacceptable, a viable solution should lead to an increased risk of not publishing the paper at hand, or reputational damage in case such behaviour is identified. As long as these risks remain low, the chances for a rigorous change might be slim. Below, a few possible solutions will be discussed with respect to the implications for risk. First, a relatively easy solution (e.g., Nosek et al., 2012; Nosek & Lakens, 2014) is to install a system of preregistration. In such a system, studies are submitted before they are conducted, and only the theoretical part and a description of the method are reviewed, and the decision whether to publish the article is conditional on the quality of these parts only. Thus, accepting or rejecting a paper does no longer depend on the results. If such a system would be implemented, there is no longer a direct need for QRPs or presenting the outcomes with certainty. It could be argued, however, that even when preregistration would be implemented, the risks of using QRPs or presenting data with certainty for hampering the publication or the researchers' reputation would remain small. It is not inconceivable that researchers would keep doing displaying those, because this was common practice for a long time, and old habits die hard. Therefore, preregistration alone may not be sufficient to unlearn such behaviour. A second solution is to require authors to make their data publically available (Wicherts, Bakker, & Molenaar, 2011). Bakker and Wicherts (2014) even claim that *not* sharing your data should be considered a QRP. If *all* the available data are shared after publishing an article, some of the QRPs are detectable, and thus the likelihood of reputational damage in case of the use of QRPs may increases. I expect that making all the data available, maybe in combination with preregistering the research questions would increase the risks of using QRPs substantially. Preregistration and making data available would, however, not necessarily prevent a too rigid interpretation of significance testing.

A third solution is to require researchers to use Bayesian instead of frequentist techniques. Although this may seem drastic, recently the journal of *Basic and Applied Psychology* published an editorial (Trafimow & Marks, 2015), in which frequentist techniques were banned from the journal, whereas Bayesian techniques are "neither required nor banned" (p. 1). Apparently, these drastic measures are implemented in practice, although this is an exception. Philosophical debates about the usability of Bayesian techniques and frequentist techniques have been going on for decades, in which both sides claim that the framework they adhere to has clear advantages over the other (see e.g., Kline, 2013; Wagenmakers, 2007) . Whether one is better than the other with respect to dealing with uncertainty is beyond the scope of this paper. I do want to stress, however, that the rigid frequentist cut-off values for significance (and it could be argued that the clear limits of CIs can be used equally rigid (Abelson, 1997)) are not intrinsically connected to Bayesian techniques. In that sense, using QRPs in order to reach a certain outcome do not make sense since there are no values that need to be reached in order to have a publishable paper. Jeffreys (1961) did propose some interpretations for Bayes factors, though, and as soon as they would be implemented as criteria for the publishability of an article, QRPs would probably be used just like they are now used with frequentist techniques. With respect to acknowledging uncertainty, Bayesian techniques explicitly deal with what a rational person should believe, assuming a certain starting position. Thus, pretending more certainty than is justified is taking an enormous risk, because every reader who understands the techniques at hand can see that the claim is unwarranted.

Most of the suggestions that are discussed above require rigorous changes on the level of editors, reviewers, and university management. If we want to solve the problems that are discussed, however, such systemic changes are probably inevitable. Nevertheless, the individual researcher does have a responsibility as well, and given the amount of attention for this issue it gets more and more dubious to maintain a waiting position. As Gigerenzer (2008) puts it: "It takes some measure of courage to cease playing along in this embarrassing game. This may cause friction with editors and colleagues, but in the end it will help them enter the dawn of statistical thinking" (p. 171). When researchers would take more risks, and the scientific community would take the risks researchers face into account, inferential techniques may become what they were originally designed for: Tools to reason about uncertainty, instead of tools to publish suboptimal papers.

References

Abelson, R. P. (1997). A retrospective on the significance test ban of 1999 (If there were no significance tests, they would be invented). In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 117-141). Mahwah, NJ: Erlbaum

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., et al. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108-119.

Bakker, M., & Wicherts, J. M. (2014). Outlier removal and the relation with reporting errors and quality of psychological research. *PloS One, 9*, e103360.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543-554.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*, 389-396.

Brooks, S. P. (2003). Bayesian computation: a statistical revolution. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences, 361*, 2681-2697.

Buchanan, M. (2007). Statistics: Conviction by numbers. *Nature*, *445*, 254-255.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*, 997-1003.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY, Routledge.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical*
inference. Basingstoke, England: Palgrave Macmillan.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274-290.

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology, 5*, 75-98.

Fidler, F., & Loftus, G. R. (2009). Why figures with error bars should replace p values: Some conceptual arguments and empirical demonstrations. *Zeitschrift Für Psychologie/Journal of Psychology*, *217*, 27-37.

Fisher, R. A. (1925). *Statistical methods for research workers.* (11th ed. rev.). Edinburgh: Oliver and Boyd.

Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology, 103*, 933.

Gigerenzer, G., 1993. The superego, the ego, and the id in statistical reasoning. In: Keren, G., Lewis, C. (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311–339). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. New York: Oxford University Press.

Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counselling for low-risk clients. *AIDS Care, 10*, 197-211.

Haller, H., & Kraus, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7*, 1-20.

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PloS One, 8*, e72467.

Hoekstra, R., Finch, S., Kiers, H. A. L., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of p values. *Psychonomic Bulletin & Review*, *13*, 1033-1037.

Hoekstra, R., Johnson, A., & Kiers, H. A. L. (2012). Confidence intervals make a difference: Effects of showing confidence intervals on inferential reasoning. *Educational and Psychological Measurement, 72*, 1039-1052.

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*, 1157-1164.

Hubbard, R., & Ryan, P. A. (2000). The historical growth of statistical significance testing in psychology—And its future prospects. *Educational and Psychological Measurement*, *60*, 661-681.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524-532.

Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, *1*, 11-27.

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences*. Washington,
DC: APA Books.

Lecoutre, M., Poitevineau, J., & Lecoutre, B. (2003). Even statisticians are not immune to misinterpretations of Null Hypothesis Tests. *International Journal of Psychology, 38*, 37-45.

Manktelow, K. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment and decision making.* Hove: Psychology Press.

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*, 137-141.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615-631.

Oakes, M. W. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Chichester: Wiley.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657-660.

Pashler, H., & Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science A Crisis of Confidence? *Perspectives on Psychological Science*, *7*, 528-530.

Scheutz, F., Andersen, B., & Wulff, H. R. (1988). What do dentists know about statistics? *European Journal of Oral Sciences*, *96*, 281-287.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359-1366.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *The American Statistician*, *49*, 108-112.

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*, 1-2.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779-804.

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS One*, *6*, e26828.

Wulff, H. R., Andersen, B., Brandenhoff, P., & Guttler, F. (1987). What do doctors know about statistics? *Statistics in Medicine*, *6*, 3-10.

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference A User's Guide. *Journal of Management*, *41*, 390-320. doi: 10.1177/0149206313501200.

Risk and Decision Making: The "Logic" of Probability

Manfred Borovcnik

Alpen-Adria University, Klagenfurt, Austria

Abstract: Risk is a hot topic. There is an international trend to use examples of risk or the concept of risk in the early teaching of probability. It enriches the problems, it widens the contexts, and it motivates the students to learn probability. This paper illustrates the notion of risk as a multi-faceted concept. The diverse perceptions start with language where risk is used in very different ways. The overlap of risk and hazard is not restricted to the technical context of safety and reliability; Knight's seminal work on risk and uncertainty has its definite impact on today's perception of the notions. The endeavour to re-interpret issues of statistical inference by risk – the risk of type I and II errors – or the concept of the weighted impact of decisions (in decision theory and in Bayesian framework) can clarify what risk means within mathematics but – as the whole machinery of statistical inference is difficult to understand – it may have little consequence on how the majority of people act and understand the notion of risk. Kahneman and Tversky show that the perception of risk is influenced by psychological factors and assert that people are risk averse in winning situations while they are risk seeking in losing situations. The perception of risk is dominated by the impact (loss or win) so that even a thorough judgement of the underlying probabilities is biased. If risk is shared between several stakeholders, they all have to use their own ingredients for their model of the same situation and follow their own logic. This leads to non-unique answers, which is unusual in mathematics. Methods of simplifying problems, the way to find a solution, and understand the underlying concepts more easily may induce a shift from a refined perception of the (hypothetical) models involved towards factual knowledge. The article aims to clarify these issues, which influence the ways to conceptualize, perceive, and teach the notion of risk.

Keywords: risk, uncertainty, risk perception, decision making, statistical errors, Bayesian risk, minimax principle.

Introduction

The reader might think that risk has a well-defined meaning and can be taught straightforwardly. The following discussion illustrates how risk can enrich probability teaching but reminds us that deliberate learning paths have to be designed so that teaching can enhance the ideas of risk in the learners. Modelling requires a different "logic" adapted to the problem being modelled and its context, which seems surprising for many as they do not perceive such diversity in mathematics and are also surprised that probability does not follow logical rules such as transitivity. In an earlier paper on risk in health we found it necessary to define risk right at the beginning of the essay:

> "By risk we understand a situation with inherent uncertainty about the (future) outcomes, which is related to impact (cost, damage, or benefit). Sometimes expected value is used for comparing several decisions, which are 'at stake'. Risk is used heterogeneously, some refer only to the probability inherent to one adverse outcome without regarding its impact, and others refer only to the adverse outcome. A decision between several choices of action might involve one person, or a decision might be 'shared' between two or more stakeholders, e.g., a patient and a doctor who have to find a decision about the next steps. […] Risk involves two components and both

are prone to subjective interpretations: the judgment of impact is different for different individuals and is even more distinct for a person or an institution with a role different from the patient." (Borovcnik, & Kapadia, 2011a, p.1)

However, one reviewer criticized our usage of the term "risk" and stated that by risk only the adverse future outcome can be meant. This would also be reflected by sentences like "The operation bears the risk of …", "I do not climb in the mountains as the risk of an accident is ...", "No risk – no fun". This shows that there is no unique concept of risk and we became aware that there are other feasible ways to define and think about risk instead of a precise definition.

Therefore, we start with a discussion of various meanings of risk and also refer to concepts linked to risk and describe situations that involve risk. In the second section we identify statistical notions of risk, which include type I and type II errors of statistical tests and Bayesian risk. The third section is devoted to psychological aspects of risk and perception of risk. We continue with three paradigmatic examples of risk that develop genuine features of risk and enrich the notion of probability by their context. In a further section, elementary approaches to probability are intended to make sense of risk involved in specific situations. A special view on the stakeholders that are possibly brought together in decision making corroborates our view that there are many ways to model and describe risk, and they all yield feasible descriptions of the situations involved but may differ in character and in what is regarded as the best decision. Implications for teaching complete the multifaceted character of risk that we project by our considerations.

Meanings of Risk

In this section, we deal with various ways to define risk, which is unusual from a mathematical perspective, where definitions are usually agreed. Furthermore, the meaning of risk – regardless of the specific definition – is also influenced by the character of probability information that is used for determining risk. We also go into details about relations between risk and hazard, between risk and uncertainty, and risk and utility. This section concludes with an analysis of situations that involve risk. The constituent parts of such situations (like who takes a risk) influence the interpretation of the input information as well as the meaning of the calculated risk.

Attempts to Define Risk

Apart from the vague use of "risk" in everyday language, risk is perceived and defined also in the scientific context. The common element between science and everyday seems to be an unwanted event that may or may not occur, often related to a low probability and severe consequences. Hansson (2007) discriminates five definitions of risk:

(1) "an *unwanted event* which may or may not occur";

(2) "the *cause* of an unwanted event which may or may not occur";

(3) "the *probability* of an unwanted event which may or may not occur";

(4) "the statistical *expectation value* of an unwanted event which may or may not occur";

(5) "the fact that a decision is made under conditions of *known* [rather than unknown] *probabilities"*.

An early definition of "risk" in the sense of (5) has been attempted by Knight (1921) who made the term very popular; his use has its roots in economic settings (see below). Generally, the use of probability, which is also controversial, is essential for the perception and interpretation of risk. Indeed, we refer to the three main classifications of probability in Borovcnik and Kapadia (2014), which are useful to clarify discussions on probability-related issues. We argue that all three approaches should be introduced in teaching probability. We will use this helpful classification throughout the paper and hope that others involved in teaching probability will use this terminology.

APT: A priori theory – probability is identified by equal probabilities in a finite probability space (Laplace theory).

FQT: Frequentist theory – probability is linked (determined, defined) by the limit of relative frequencies.

SJT: Subjectivist theory – probability is the personal degree of belief of a person in an uncertain statement.

There is an axiomatic justification for each approach and its inherent interpretation. The three approaches have led to fierce controversies and disputes historically. This is partly because there are restrictions for each of them. There are no situations with equal probabilities, there is no limit of relative frequencies in the finite series of experiments in the real world, and the "elicitation of probabilities" is too subjective for many people. Besides and along all these approaches, probability is also used as a scenario term to investigate a real problem on the basis "what happens if ..." (see Borovcnik, 2006; 2011). An example of the use of the scenario idea is the following:

How many redundant technical units (built in a system in parallel so that all have to fail for the system to fail) are required to attain an acceptable risk (probability) of failure of the system *if* the reliability (probability of functioning) of single units is 0.95 and the failures are independent? Mathematically, one can calculate the system's reliability as $1-0.05^n$ and compare the increase in reliability to the increase in cost and decide about a reasonable balancing point. The values calculated allow for a transparent decision but are not to be taken literally as approximating any realistic probability (neither as property of single units nor of the system).

The meaning (4) about risk is the one preferred in technical environments dealing with the reliability of systems. It seems more objective and thus widely communicable or compulsive. However, its relevance still depends on the actual probabilities involved. There might be limited documentation about failures of units, there might be only qualitative statements of engineers (from stress tests under conditions of highly elevated stress levels, which have an effect on shorter life time and earlier failures), and there might be only general considerations of redundancy as in the scenario above. In cases when there is a very low probability of failure of a unit (which is often true), there are no failures to observe in real life under normal conditions so that any probability is a scenario figure.

Though the discussion of risk has been initiated by Knight in an economic setting, the technical usage of risk has been made popular by early studies on nuclear reactor safety (Rasmussen et al., 1975). It is important to regard the character of probabilities involved in risks that are communicated as well as the precision of knowledge on the probabilities used to calculate the risk. To speak in terms of risk about probabilistic situations is just a twist in language. Similar problems have been dealt with since the early beginnings of probability. And they have led to unsolvable paradoxes like the St Petersburg paradox, in which the fair stakes of a bet are derived as an infinite amount (see Borovcnik, & Kapadia, 2014). This also prompted considerations of the utility of money for the first time, by which the expected value of the game turns to the risk in the sense of (4). The change of terms, however, reflects also a change of applications of probability, especially to the economy (where risk is an inherent component of actions) and reliability and safety considerations in technology (where risk is also shared between stakeholders of unequal access to influencing the decision).

Risk and Hazard

Various dictionaries (like Merriam-Webster) also link the adverse event with the potential underlying cause and subsume it under the heading of "risk", such as the possibility that something bad or unpleasant (such as an injury or a loss) will happen, someone or something that may cause something bad or unpleasant to happen, or a person or thing that someone judges to be a good or bad choice for insurance, or a loan.

There are key differences in the five definitions in the previous section. In (1) there is an unwanted event; in (3) it is the probability of an adverse event; a summary figure for risk occurs in (4); in (2), risk is *something that may cause an adverse event*. There is a subtle difference of wording in (1) and (2) with the latter focusing on the factors behind, which might cause the harm. Such factors are called *hazards*. However, usage of the notion hazard is diverse outside technical contexts. Often dictionaries do not give specific definitions of it or combine it with the term "risk". For example, one dictionary defines hazard as "a danger or risk" or refers risk back to hazard which helps explain why many people use all the terms interchangeably and contribute to the confusion. Relating to workplaces, CCOHS (n.d.) states: "A hazard is any source of potential damage, harm or adverse health effects on something or someone under certain conditions at work." Examples of hazards at the workplace are materials like asbestos that may cause mesothelioma, or substances like benzene that may cause leukaemia.

In this setting, hazard becomes what earlier has been subsumed in (2) as risk, and risk becomes the chance or probability that a person will be harmed if exposed to the underlying hazard; this is (3), not (4). Risk assessment becomes the process of identifying hazards and evaluating the risk associated to the specific hazard; risk management is then any measurement to eliminate or control hazards at an acceptable level.

The problem with hazards is that they are linked via probabilities or chance to the adverse events. If the effect is acute (with immediate consequences) the link back to the hazard is obvious. However, more often the effect is chronic and delayed so that a link is hard to establish and it is difficult to provide evidence for. Potential hazards are statistically identified by correlations or associations (ex post exposure rates) and rarely can a causal link be established. That makes it very difficult to convince people about (potential) hazards.

When we study the relationship between smoking and the occurrence of lung cancer, we notice that the percentage of people with lung cancer is greater in those who have smoked earlier; that *indicates* that it is better not to smoke in the sense of doing something in order not get lung cancer. However, this association cannot directly be interpreted in *causal* terms as we cannot exclude third factors (hidden variables) operating in the background. One factor may be problems in handling stress that can lead people to smoke and at the same time make them more prone to lung cancer.

Another, simpler example illustrates the dilemma. If in the world-ski championship the Austrian skiers win one third of the medals, then we cannot infer from this fact that one third of the Austrian skiers are serious candidates for medals. They are not all the same, which is the basic assumption in the cancer example from before. They differ by third variables that might affect their success (disease above). This illustrates that cause and effect cannot simply be exchanged. Statistical methods are specially designed to corroborate statistical evidence; they indicate in which cases it may be rewarding to search for contextual interrelationships but cannot provide a substitute for such research.

Knight's Distinction of Risk from Uncertainty

According to Knight "risk" and "uncertainty" differ by the character of probability: if the underlying probabilities are not known, then Knight speaks of "decisions under uncertainty", if they are known then he calls them decisions of "risk". This distinction clearly implies that risk excludes uncertainty, i.e., once the probabilities (and the impact) for the various events are estimated, uncertainty (or lack of any knowledge) is eliminated from the situation. Knight's terminology has been so successful as – on the surface – it shifts the connotation of risk away from uncertainty (in the general sense) and makes the term "risk" seemingly more objective and thus more acceptable, which may be advantageous. The original quotation is:

> The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; […] It will appear that a *measurable* uncertainty, or 'risk' proper […] is so far different from an *unmeasurable* one that it is not in effect an uncertainty at all. We shall accordingly restrict the term 'uncertainty' to cases of the non-quantitative type. (Knight, 1921, pp.19)

However, there are several objections. The use of "uncertainty" by Knight conflicts with its use in probability and philosophy. Philosophically, uncertainty (or chance) is opposed to determinism. Uncertainty is the general term designating (random) situations where a precise prediction of an outcome cannot be made with certainty, i.e., where probability has to be used in the prediction. Whether or not it is possible to get good estimates or relevant hypothetical values of the probabilities involved does not matter as the character of the probabilities (APT, FQT, or SJT) is not relevant. As the future is always uncertain, a risk is attached to actions relating to it.

A second objection is that Knight adopts a very naive usage of probability as – apart from textbook examples – probabilities are *not* known: even in games of chances there is always the dispute whether the conditions of a Laplace experiment with equal likelihood are in fact fulfilled in a specific case. Two other positions can also be stated: either (following von Mises) that probabilities are always only partially known and have to be estimated (FQT), and, probability does not exist in the objective world and is only part of our judgement on the world (SJT), as espoused by de Finetti. What we know is that the numbers are related to the models (e.g., an equiprobability model for a die yields 1/6 for each of the faces in the sense of APT probability) or estimates from past experience (FQT probability). Or, we know the values from elicitation from a person's preference system (SJT probability).

A third objection is that from a subjectivist viewpoint the consistent choices of a rational person yield his or her subjective probabilities. Thus, while we have no precise values for probabilities (in FQT sense) we have probabilities (in SJT sense) and therefore the distinction between risk and uncertainty is not valid as noted by Friedman (1976, p.282).

Uncertainty has to be met by suitable procedures to get estimates or qualitative knowledge on the unknown probabilities. Of course, the status and the precision of probabilities that are used to calculate risks in the sense of (4) heavily influences the possible interpretation of this risk and its relevance for a discussion, especially if it relates to different stakeholders as is in the public discussion of safety of nuclear power or in public health issues. Of course, the more frequentist information about the underlying probabilities is available, the more objective seems to be the calculated risk. However, still there might be a dispute as different stakeholders can (and have to) exploit different information for their purposes and they also "suffer" and benefit differently from the various actions to avoid the inherent risk (see Borovcnik, & Kapadia, 2011b).

Today, researchers defend the Knightian distinction between risk and uncertainty by saying that Knight did in fact link uncertainty to situations where markets collapse and so no subjective probabilities can be found (LeRoy, & Singell, 1987).

Knight was a key person in founding the Chicago School of Economics (with 3 Nobel Prize awards) and his ideas about risk and uncertainty have influenced many people. One example may illustrate how they express their ideas in Knight's tradition:

> "*Risk*: We don't know what is going to happen next, but we do know *what the distribution looks like*. *Uncertainty*: We don't know what is going to happen next, and *we do not know what the possible distribution* looks like. In other words […] the future is always unknown — but that does not make it 'uncertain'." (Ritholtz, n.d.)

This view on the financial markets from a former trader, researcher and strategist, and now asset manager reflects a desire of eliminating uncertainty and arrive at objective figures for probabilities: there is risk but *no* uncertainty, and we have everything under control. This wording induces a misleading impression of the measurements taken or possible and reflects the fact that "experts" adopt a terminology that fits their purpose (rather than suiting their client's interests).

Risk and Utility in Economic Theory

Risk is a key factor for economic activities and is the basic element for decisions on financial markets, which resembles the old paradigm of bets. In modelling economic activities, risk plays a central role therefore. If actions with the same expected utility are compared according to (4), the one with the larger variance is defined as more risky. Risky behaviour can then be described by properties of the utility function; to be risk averse means to prefer that one of two actions (with the same expected utility), which has less variance (in utility). The degree of risk aversion may be measured by the person's willingness to pay to get one option over the other that is offered. The Arrow-Pratt measure of risk aversion is defined as $-u''(x)/u'(x)$ ($u(x)$ the utility function, which is presumed to be twice differentiable); the higher this value the more risk averse is the person.

Apart from theoretical frames like expected utility as the criterion for decisions, researchers also became interested in how people behave in comparison to normative theories and explain – mainly by reference to psychology – why they deviate from normative results. Kahneman and Tversky are famous for their experiments, in which they investigated how people really behave in simplified situations. People seem to replace utility by an asymmetric function and put much more weight on the tails of the probability scale, i.e., differences in probability have much more influence on actions when they relate to very small and very large probabilities as compared to probabilities in the middle (around 0.50). Tversky and Kahneman (1986) developed the so-called prospect theory to describe how actual behaviour deviates from theoretical models of rational behaviour under risk (for their experiments see below).

Analytic Investigation of Situations that Involve Risk

So far we have reviewed different concepts (1) to (5) to define risk and to discern risk from neighbouring concepts. The adverse events bear subjective components in two ways: their probabilities are not always judged by objective (FQT) information and the evaluation of impact may be biased by personal circumstances. A fuller comprehension of the mathematical concepts as well as psychological factors may play a key role in processing the available information (in the extreme case even apparently objective information is just ignored or reinterpreted). We will deal with these issues separately below.

From a systems analytic approach, it does not suffice to define risk but it is essential to structure those situations that involve risk: which constituents build up a risky situation and how do they influence the way to deal with risk or the perception of risk? Borovcnik and Kapadia (2011b) refer to the following ingredients of a decision situation related to health: "the nature of the risk (and mathematical concepts like probability); the psychological matters involved; the type of situation (treatment, prevention) and information used; the people involved, their aims (purpose) and their inherent criteria". These are general categories to describe risky situations.

Risk may involve only the decision maker who takes personal risks like travelling, smoking, nutrition, or personal health issues. A decision between alternatives is influenced by a subjective judgement of a balance between benefits and adverse effects. Decisions may also bring two or more stakeholders together, which typically is the case in (public) health. It is important to note that the stakeholders do not use the same information, or the same criteria to process the information, and they are also affected quite differently by the outcome. While a patient personally might suffer from consequences (no cure, side effects of treatment), the doctor is liable to the system to guarantee treatment according to the state of art. Other stakeholders in public health (including the media) still experience other impact and follow other optimizing criteria; for the dilemma resulting from the interests of the various stakeholders see Borovcnik and Kapadia (2011a, b). It is clear that the different stakeholders have to use different criteria to find or justify their decisions. Remarkable, however, is the inconsistent use of criteria by individual decision makers (see below).

The type of the situation is a further, often neglected influence factor on the perception of risk: there are "real" risks when an acute situation is there and some action has to be undertaken; think of an accident with a broken ankle. While there are several options, something has to be done to deal with the immediate situation. Sometimes there are virtual risks relating to the (far) future. The person might conceive cancer of the prostate. There are hazards lurking in the background but no direct link can be made to the adverse event. In public health, screening programmes have been installed for various diseases: mammography with breast cancer (see Gigerenzer, 2002, for general risk considerations in this case); or, PSA tests and biopsies with prostate cancer (see Sandblom, et al, 2011, for an attempt to evaluate the success of the programme). Another virtual risk is climate change. Whether or not there is a climate change, whether or not such a change will have adverse effects directly related to it, there is much effort and investment of money in an attempt to decrease the "identified" hazards for a climate change.

Statistical Notions of Risk

In this section we present statistical terms that – implicitly – deal with risk, or, vice versa, that build up a normative way to conceptualize risk. Risk is linked to the possibility and probability that a statistical test leads to a wrong "decision" and is identified as conditional probability. The Bayesian approach to conceptualize risk is linked to an expected loss argument. We also discuss the problems that arise from low probabilities as we have no data for such situations so that the decisions (and related risks) based on them lack empirical justifications.

Type I and Type II Errors

In the historical development of statistical tests, Neyman and Pearson used a decision theoretic framework. In the simplest case we have two probability distributions for a phenomenon and we do not know, which distribution really applies. We decide in favour of one after we see the data from a random sample from the "true" distribution.

Example from quality control. A production line fabricates items that have an average rate of defectives of p = 0.04 if the process is under control. If there is a disruption in the process, the proportion of defectives increases. We restrict the possible values from $p > 0.04$ to $p_1 = 0.10$; i.e., 10 % defectives are produced on average in this case. We have now two hypotheses; the null hypothesis H_0: $p_0 = 0.04$ and the alternative hypothesis H_1: $p_1 = 0.10$ and we search for a reasonable decision rule if we inspect a sample of n = 100 items from the current production. The situation is represented in Table 1.

Status	Decision	
	T_0: do not reject ("accept") H_0	T_1: reject H_0
H_0: $p_0 = 0.04$ Production is under control	Decision is correct	Type I error H_0 is erroneously rejected though it is true The production has been stopped yet everything was okay
H_1: $p_1 = 0.10$ Production is out of control	Type II error H_0 is not rejected even though H_1 is true A serious disturbance of the production has not been detected	Decision is correct

Table 1. Risk of wrong decisions in a statistical test of the null hypothesis H_0 against an alternative H_1

There is a risk of wrong decisions in the sense of (1) as Neyman and Pearson formulated in their papers. The probabilities of wrong decisions (the adverse event) are conditional probabilities (conditional to the distributions in the null and alternative hypotheses). These probabilities of wrong decisions are often denoted by α and β and can be called risks in the sense of definition (3): $P(\mathrm{T}_1 \mid \mathrm{H}_0 : p_0 = 0.04) = \alpha$ and $P(\mathrm{T}_0 \mid \mathrm{H}_1 : p_1 = 0.10) = \beta$. The decisions T_0 and T_1 are based on the number X of defectives in the sample that is inspected. By the usual assumptions, X is binomially distributed with n (=100 here) and $p_0 = 0.04$ if H_0 and $p_1 = 0.10$ if H_1 is true. From here we can calculate α and β if we use a threshold of X = 8 for our decision, i.e., if we reject H_0 if $X \geq 8$ (T_1) and do not reject H_0 if $X \leq 7$ (T_0).

$P(X \geq 8 \mid p_0 = 0.04) = 0.0475$ and $P(X \leq 7 \mid p_1 = 0.10) = 0.2061$. Usually, the type I error is pre-set (often to 0.05. or 0.01) so that the threshold can be calculated and the type II error is dependent on the size of the type I error. In this setting we cannot obtain probabilities for wrong decisions, which are $P(H_0 \mid T_1)$ and $P(H_1 \mid T_0)$. In FQT view, hypotheses cannot have probabilities as there is no experiment for them (and therefore no relative frequencies); in SJT view, hypotheses have probabilities as any (unknown) statement or event does have a probability but as long as the probabilities are not "elicited" the probabilities of wrong decisions are not known.

Errors of type I and II are interpreted by the following thought experiment (scenario; under laboratory conditions): If the null hypothesis always applies, then it will be rejected with probability α, that is, in $100\,\alpha$ % of cases applying the test, the null hypothesis will erroneously be rejected. For our production line example this means that for $\alpha = 0.05$ the production will be stopped in 5% (on average) to search for specific causes but the production definitely is under control. If β = 0.20, then the production would not be stopped though the conditions of the alternative hypothesis apply. For the production line this means that – if the production is out of control and items are produced with an average rate of defectives of p_1 = 10 % – the decision rule would not "recognize" that with probability 0.20 and let the production continue.

Such a primitive FQT interpretation is misleading as one never tests under laboratory conditions and it provokes an interpretation of α as an error rate that conveys something on the risk of wrong decisions. As we know that it is not an unconditional probability, we change the names and refer to the risk rather than the probability of a type I error. Taken literally, the pure FQT position leads to rationality gaps, which have been fiercely criticized in the controversy about the foundations of statistics (see Stegmüller, 1973, or Hacking, 1975, 1990). Thus, the type I error should be interpreted as qualitative figure that is needed to derive the threshold and the type II error is just a mathematical consequence of it. If it seems too large (as here), then more data should be sampled before the decision is made. The Bayesian approach would compare various decision rules according to expected loss in the sense of (4). We will see the approach at work in the example on journal copies later.

Fisher suggested to test a null hypothesis against data from a random sample and insisted on not to formulate a specific alternative hypothesis. Instead, he based his decision of rejecting the null hypothesis on the *p*-value, which is the probability of getting a sample statistic (say, a sample mean) as far as, or further (more extreme) from some reference value (which reflects the true population parameter). Fisher thought that the *p*-value could be interpreted as a discrepancy measure from the null hypothesis. Thus, the *p*-value would be interpreted as error probability of the test. However, as we have seen in our framework above, it reflects only the probability of erroneously rejecting a true null hypothesis, i.e., $P(\mathrm{T}_1 \mid \mathrm{H}_0)$ and not the probability of an error in decision making if we reject the null hypothesis, which is $P(\mathrm{H}_0 \mid \mathrm{T}_1)$. Thus it is misleading to associate the risk of a wrong decision with the *p*-value.

In the early attempts to formulate his ideas about significance tests, Fisher pre-sets the significance level α and rejects the null hypothesis if its *p*-value is lower than α. In comparison to the Neyman-Pearson test policy, this significance level coincides with the type I error. However, the significance test lacks information about specific alternatives and their related type II errors (β). The complement $1-\beta$ is also known as the power of a test and conveys information about the probability that the test can "detect" a specific deviation from the null hypothesis if such a deviating hypothesis would actually be the true distribution for the investigated phenomenon. The power is a key concept to judge the risks associated to a statistical test.

While there are strong arguments for Neyman-Pearson's view, in practice an alternative can often not be formulated like in the ANOVA case, when one has not yet a clear idea about the structure of the differences in mean between the tested populations and it is methodologically doubtful to extract hypotheses from the same data, which are used to test these hypotheses; or, in testing for independence, where it is difficult to formulate specific hypotheses for dependence. Yet, Lehman (1993) tried to reconcile both approaches towards testing.

Risks with Low Probabilities

For events with very low probabilities, we have no data available and thus any FQT consideration is doubtful. What remains is an elicitation of subjective probabilities or the use probabilities in a scenario to investigate the issue on a "what if" basis. We illustrate the difficulty of getting reliable data for low probabilities by an example from Borovcnik (2012, p.22). Normally, an unknown probability is estimated by the related relative frequencies from a sample that is assumed to be random. This latter assumption is hard to meet, hardly questioned, and really difficult to check (a test for randomness lacks an alternative and therefore, a type II error cannot be calculated). Such an estimate has an inherent variability. Suppose, we want to investigate how we can get information for a probability p of 10^{-4}. If we base an estimate of p on a sample of "10,000 there is a 36.8% chance to get no one with the property with an estimate of $\hat{p}=0$, a further 36.8% for an estimate of $\hat{p}=0.0001$ (exact), 18.4% for an estimate of $\hat{p}=0.0002$ (doubles the value), 6.1% for $\hat{p}=0.0003$ (triples it) and 1.9% for an estimate more than four times the underlying value".

Another example is bovine spongiform encephalopathy (BSE, or mad cow disease), taken from Dubben and Beck-Bornholdt (2010, pp.64). There were times when all of Europe was in uproar because of news on new "cases" of BSE. Cattle with the disease are a hazard as those who eat their meat may contract Jacob-Creutzfeldt disease. Tests have been undertaken to evaluate the accuracy of diagnosing whether an animal has or does not have the disease. With the diagnosis test applied, a cattle gets the attribute positive T^+ (indicating the disease) or negative T^- (indicating that the animal is free of this disease). If the cattle has BSE, there is a "sensitivity" of (at least 0.99) that it will get a positive test result; if the cattle has no BSE, there is a "specifity" of (at least) 0.997 that it will get a negative test result. Both sensitivity and specifity are conditional probabilities: sensitivity $= P(T^+ \mid \text{BSE})$; specifity = $P(T^- \mid \text{no BSE})$.

These properties of the diagnosing procedure have been based on testing 300 infected cattle from British origin and from 1,000 animals from disease-free New Zealand. And in German laboratories all 1,300 biometric samples have been classified correctly, from where the properties above have been calculated by means of confidence intervals. However, put it the other way round and calculate the probability that 1,000 non-infected cattle are classified correctly as T^- given that the specifity is 0.997: to classify them all negative, there is – merely by chance – a probability of 5%. The extraordinary test result may thus be explained by mere randomness.

This reduces the 100% security of the laboratory test considerably. And, furthermore, remember that it is much more difficult to detect the still latent disease when no apparent traces of it are there than it is to confirm the disease (by the same diagnosing test) of a full-blown BSE cattle. That is, under conditions of mass application the careful procedure of the laboratory test may not be upheld. A cattle classified positive need not be infected; we cannot claim that with certainty. Likewise, a cattle classified negative need not be disease-free. To calculate the probability that a cattle with diagnosis T^+ in fact has BSE, we need also the prevalence of BSE in the population of German cattle. This is an unknown number. While there are subjective estimates, there is no objective number.

There were 331 cases of positive tests in Germany from January 2001 till June 2004 when 9,747,738 cattle have been tested. That means the prevalence is less than 3 out of 100,000. It could well be that this number is much lower and all the positive test results have been false positives. We will never know. This illustrates drastically the problem with hazards that have a low probability to cause harm. As the impact is high (human beings may contract a deadly disease), societal rationality is put under enormous stress.

Another issue arising from low probabilities is that we underestimate cumulative risk. Suppose there is a personal risk in climbing. For one trip a risk might be calculated in the sense of (3) of 0.0001, which is negligible. For 2,000 trips the probability of not being involved in the adverse event attached to it is calculated to $0.9999^{2,000}$ whence the cumulative risk is 18%. What if the initial risk is 0.0003 (which is still a negligible number)? The cumulative risk rises to 45%. Who would have expected that effect? The difficulty in understanding such calculations means that the cumulative risk is neglected though it is essential.

Psychological Aspects of Risk

Risk involves the probabilities of several outcomes and their impact. The probabilities may attain a more subjective (SJT) or a stronger objective (FQT) character. The impact can be measured by "money" or utility; the latter has a strong subjective component. It is interesting to know whether people share similar forms of utility and how far they are consistent between different choices they make. A further key issue is whether people behave rationally or are influenced by psychological factors. The criteria people choose might be dependent on their perception of risk, which may be influenced by context, by low probabilities, or by high impact. Surprisingly, there are arguments to change the criterion for decisions depending on whether a decision is made once or made repeatedly. We start with some famous experiments of Kahneman and Tversky.

Probability and Utility

Kahneman and Tversky conducted many experiments with adults on understanding probability and risk and showed how inconsistent people are with regards to uncertainty. They characterised their behaviour as misconceptions and formulated a terminology: representativeness, availability, anchoring, etc. The examples used to explore people's behaviour resemble the following one though some are more complicated (Kahneman, & Tversky, 1979; Tversky, & Kahneman, 1981).

Experiment 1: Win [in $]	Option a_1	Option a_2
Potential future	1,000	2,500 with ½ 0 with ½
Experiment 2: Win [in $]	Option a_3	Option a_4
Potential future	−1,000	−2,500 with ½ 0 with ½

Table 2. Type of experiments, Kahneman and Tversky (1979) used for their studies

In experiment 1 (see Table 2) people prefer $ 1,000 for sure even if the second option has a value of 1,250 while in experiment 2 people prefer the second option and seek the risk rather than the sure loss of $ −1,000. In winning situations people are risk averse. People prefer $ 1,000 for sure even if the second option has a value of 1,250. On the contrary, in losing situations people are risk seeking: they prefer the second option with a value of −1,250 and seek the risk rather than the sure loss of −1,000.

> "Problems […] demonstrate that outcomes which are obtained with certainty are overweighed relative to uncertain outcomes. In the positive domain, the certainty effect contributes to a risk averse preference for a sure gain over a larger gain that is merely probable. In the negative domain, the same effect leads to a risk seeking preference for a loss that is merely probable over a smaller loss that is certain. The same psychological principle – the overweighting of certainty – favours risk aversion in the domain of gains and risk seeking in the domain of losses." (Kahneman, & Tversky, 1979, pp. 268)

In developing prospect theory Kahneman and Tversky (1979) found that the decisions observed in their experiments are incompatible with expected utility theory that has been proposed as normative theory of behaviour in view of risk by Neumann and Morgenstern (1953) and widely used as a model to describe economic behaviour (Arrow, 1971). In 2002, Kahneman was awarded a Nobel Prize for his work on the psychology of decision making (of which prospect theory is a foundational part).

With the usual utility functions (concave for positive and convex for negative values of money), such behaviour is inconsistent. That implies also that people have to act in the experiment as if it were real which is doubtful as in virtual games decisions are different from the hard consequences of real life where no switch-off button exists. The method also requires that people can disentangle the perception of probabilities and utility considerations; however, the perception of any probability is always linked to its impact (the probability of an event with a severe negative impact is often highly over-estimated). Furthermore, Kahneman and Tversky only described the phenomenon that has since become famous: people are risk-averse in winning situations and are risk-seeking in losing situations. They did not explain this behaviour by their analyses.

We will twist the representation of the experiments by a reformulation, which changes the character of the winning situation to a losing situation and vice versa. By this we show that the representation of a situation is not unique and implications of the behaviour cannot easily be drawn. Furthermore, we will explain the behaviour, which seems obvious in the new format. And, the result can be reconciled with utility with a somewhat different utility function, which has not yet been considered but seems relevant (one that is bounded from below and the lower bound is reached very "soon") especially if one investigates ongoing political and financial decisions around the rescue of the Euro since 2008.

Experiment 1 can be formulated differently. People think of \$ 1,000 for sure as they have it already with option a_1. The decision for a_2 then appears as relative to the status quo as potential losing situation as they imagine losing \$ 1,000. Instead of comparing a_1 and a_2, they compare a_1 and a_2^*. This sheds doubt on Kahneman and Tversky's hypothesis "in winning situations people are risk averse" (Table 3) and "in losing situations people are risk seeking" (Table 4).

Experiment 1^*	Option a_1	Option a_2^*	Option a_2
Potential future	1,000	1,500 with ½ –1,000 with ½	2,500 with ½ 0 with ½

Table 3. Reformulation of experiment 1 from a winning to a losing situation

The possibility of winning an extra amount of only \$ 1,500 is now related to the additional risk to lose the 1,000 and go down to zero! It may not pay to risk the 1,000 for the small extra win of 1,500. The situation may thus be perceived as losing situation. A person that has already 1,000 has to be paid (much) more to take a risk to lose 1,000. The more a person already owns, the more this person has to be paid to seek the risk.

Experiment 2*	Option a_3	Option a_4^*	Option a_4
Potential future	−1,000	−1,500 with ½ 1,000 with ½	−2,500 with ½ 0 with ½

Table 4. Reformulation of experiment 2 from a losing to a winning situation

In the same way we can reformulate the situation of the second experiment and regard the option relative to the status of having already lost 1,000. The situation is now a winning situation. People who already have debts will take more risks to balance them. In our own experiments with students for these tables, we have found that the hypothesis of Kahneman and Tversky is not supported; furthermore, we have also found that people are markedly influenced by their own disposable income.

Inconsistencies and Different Perceptions

Behaviour may be influenced if utility of money is involved, if payments are much bigger, or if initial situations are perceived (presented) differently. It might be the same people who avoid a_2^* and seek a_4^*, i.e., avoid winning an extra 250 (as expected value) but seek to lose another 250. It is not inconsistent as it compares the behaviour of the same person in two different settings (one with already a reasonable amount of money, the other with an unbearable loss). That changes the alternatives at stake and influences their behaviour switching their decision. It is important to note that judgements can only be seen with the related alternatives at stake. Thus, it is not feasible to compare a_2^* and a_4^*directly.

One can now change the alternative involved to trace the consequences thereof. If we consider the offer to win a very high (but improbable) prize compared to a small amount to pay (as in the state lottery), we will see that a lot of people are risk-seeking in this winning (in the sense of Kahneman, & Tversky) situation. On the other hand, if a potential (but improbable) high loss can be "balanced" by a relatively small amount of money as in an insurance contract (which is very popular nowadays as people try to get a contract on nearly everything to eliminate uncertainty) in losing situations, we see evidence that they are risk-averse in losing situations (again in the Kahneman, & Tversky sense).

The two latter findings about the ubiquity of games of chance (and betting) and the popularity of insurance contracts show a distinct deviation from what Kahneman and Tversky claim and gives our reformulation more justification. Winning or losing is not a predicate for a situation that can be unambiguously granted. It is always a matter of reconstruction of the situation involved. Rather than to regard a situation as winning or losing situation, one might focus better on the original status (initial wealth) of the test person. If a person earns 1,000 per month or 5,000 it does affect the decision; it might influence it much more than the various amounts that are at stake; similarly, if a person has great debts already with no perspective to fulfil them, this person might be willing to accept risks that other persons would find completely irrational; indeed that is the basis for some gamblers. Also, these "facts" indicate that people do not focus so much on probabilities but much more on the impact related to the potential outcomes.

The preceding considerations also indicate that it is difficult to separate between the perception of impact and probability of future outcomes. Personal criteria come to the fore. In a winning situation, beliefs like "I deserve it (to win the prize and lead a happy life)" let people gamble (seek the risk) and wait for "God's decision". In a losing situation, they are willing to pay an amount of protection money to ease out "God's enragement" to avoid not only the "accident" but also the negative consequence related to it that goes beyond the financial loss. The personality is a further factor that influences decisions under uncertainty; it is possible that women – on the average – have a distinct risk profile from men. All these considerations reveal how difficult it is to weight the evidence rationally. Gigerenzer (2007) investigates the effect of gut decisions on the quality of decisions: first, gut decisions are popular; second, they are not necessarily worse than careful analysis of all givens; third, it might be good to support quicker decisions by checklists about what to consider. A mixture between analysis and gut decisions might be superior to pure analytic decisions not only because of the undue delay in decision by careful analysis including collecting all the necessary information.

A final example about personal criteria refers to the three-door problem with the two goats and the car behind the doors where the candidate can freely choose one. After the candidate has made the choice, the moderator opens one of the remaining doors to show a goat and offers the option to change the first choice, i.e., to switch to the still closed door or remain with that door selected first. Borovcnik (2012) discusses various feasible strategies to analyse the problem and whether it is advantageous to change one's first choice. Of course, by switching the candidate increases the chances to win the car from 1/3 to 2/3. However, with all the explanations, people – in the majority – stubbornly remain with their first choice even if they are fully capable to understand the solution. The discussion is signified by fierce emotional expressions so that one may ask for the reason behind. One simple explanation to refute all mathematical solutions might be that the candidates simply compare the situations afterwards and relate a strong impact to these: to stay with the first choice and to lose will leave them as unlucky; they simply had not a "fair chance" to win the car; to lose the car if they had chosen the right door and to switch to the remaining door will leave them with the irony of friends and with the feeling that they should not challenge God's will. They do not compare risks in the sense of (3) or weighted risk in the sense of (4); no, they simply compare the impact of their decision. And the impact of losing by their action and being responsible for this is by far the worst they can imagine. Therefore, they defend their first choice against all odds.

Idiosyncratic criteria, gut feelings, fear of high loss in the worst case, eager, responsibility for what one does (as compared to bad luck) always overlays a rational approach; it starts with hindering to separate the impact from the (small) probability of occurrence; it biases the perception of the (small) probabilities of adverse events; and it increases the factual impact of the adverse event and focuses on the worst case letting any weighting seem irrational and unfeasible.

The Logic of Repeated Decisions and One-Off Decisions

In a statistical variant of experiments 1 and 2, a candidate has to decide repeatedly (1,000 times) between the presented options. Instead of performing the experiment, we just tell what we would do and give a rationale for it. What would the reader decide in the situations presented in Table 5? Again, utility may play a role but as the payments are so small, more or less the amount counts (this argument might not apply if the single amounts were higher). We decide for the risky option b_2 in the winning situation of experiment 1^s and for option b_3 in the losing situation of experiment 2^s. Thus we seek the risk in the winning situation and avoid the risk in the losing situation (terms applied in the sense of Kahneman, & Tversky), which is different from our behaviour in the non-statistical variant of the experiment. We will not transform the situation relative to the "left" side but give a statistical argument that should suffice.

Experiment 1^s: Win [in $]	Option b_1	Option b_2
Potential future	1	2,5 with ½ 0 with ½
Experiment 2^s: Win [in $]	Option b_3	Option b_4
Potential future	–1	–2,5 with ½ 0 with ½

Table 5. Statistical variant of experiments 1 and 2: 1,000 games have to be played; each requires a decision between the two given options

To win an amount of more than 1,000 in the first experiment, we need to win in more than 400 single situations. The probability can be easily calculated from the binomial distribution (with $n = 1{,}000$ and $p = ½$) to a figure, which is close to 1 ($1–10^{-10}$). We have a probability of at least 0.51 to win 1,247! For the losing situation now the results are just reversed: we have a probability of nearly 1 to lose more than 1,000 and we lose 1,247 with a probability of at least 0.51.

It makes a difference in terms of the success of the strategy used if one has a one-off decision or decides similar cases repeatedly. What is good in a one-off decision may be bad for the repeated decision. In other words, to speak with an illustrative context: if one sells a house one time, the relevant criteria differ from those if one sells houses every week like a professional agent; this is not because of the impersonal impact but because of doing it repeatedly; of course, a professional agent has additional criteria for decisions, too.

Three Paradigmatic Examples of Risk

In the following three examples, we present the genuine character of situations under uncertainty and discuss various useful strategies. The criteria used are not always defensible and they have their relative merits and disadvantages. Of course, the decision based on a specific criterion will depend on it and possibly switch if the criterion is changed. If probabilities are involved in the decisions, their values can change even in the same problem for the various stakeholders involved. There is no unique view of a problem; rationality depends also on the position of a stakeholder in a decision. Decisions under uncertainty – if not "played against nature" may also be interpreted as an exchange of certainty and money between the stakeholders as is typically done in an insurance contract, which is a generalization from bets within games of chance.

Example 1. Insurance – Exchange Money for Certainty

We will embed matters in a decision with two stakeholders that exploit different types of information for their decision. In any insurance contract, two partners mutually exchange money and the status of uncertainty. For a full-coverage insurance for the car (for one year), the insurance company gives up its position of certainty (no loss) and offers to pay the potential costs from an accident while the client pays a certain amount (the premium) in advance in order to leave the situation of uncertainty (of an accident) and reach a status of certainty over the financial cost of an accident. A similar situation occurs in any bet in the state lottery, bets in sports, or with option contracts in the financial market.

If both parties apply the same principles then the question arises how can the contract be advantageous for both – it seems like a paradoxical situation. However, both stakeholders have their own viewpoint and use different criteria that fit their circumstances better. The insurance company has many contracts and thus can use the frequencies of accidents from past statistics to estimate the underlying probabilities. Since the company accumulates assets, it is free of utility considerations. The client, however, is a unique person with different habits (driving skills, driving regions, exposition to special risks, etc.), which emerge into a personal judgement of probabilities of accidents. Furthermore, the client does not have large financial assets so that utility considerations of money become relevant, e.g., can one afford the cost of another new car within one year in the worst case of an accident? It is interesting that states normally have no insurance for their cars. They have a larger fleet of cars and they have a different financial background. We discuss a crude model for the potential future of one year considering only a total wreckage or no accident below (Borovcnik 2006; see Table 6).

	Cost (€)	a_1 = Insurance yes	a_2 = No insurance
Potential future	t_1 = No accident	1,000	0
	t_2 = Total wreckage	1,000	30,000

Table 6. A crude model for the insurance contract

The insurance company may base its model on money and an estimate of the probability for the damages (and related payments) by past frequencies of events. For data of 2% for the wreckage the calculation is: $30{,}000 \cdot 0.02 + 0 \cdot 0.98 = 600$ plus ... expenses and profit. The car owner has to find his or her personal probabilities and take utility of money into consideration. Without utility, a so-called break-even point can avoid the difficult process of eliciting the personal probabilities. The break-even here is for odds of 1 : 29 for the total wreckage as in this case the decisions for the insurance or against it each have the same cost, namely 1,000 so that the person (not considering utility) would be indifferent between the two actions. If the model considers also smaller accidents (parking damages, e.g.) taking out the insurance becomes more attractive.

Example 2. Using Probabilities for Optimization of a Decision under Uncertainty

We illustrate how a person willing to model uncertain future outcomes by probabilities may optimize a current decision. Suppose the demand D for a journal is uncertain. We model it – for reasons of simplicity – by the following discrete probabilities p_i. Cost of production $C(a_j)$ (in €) is dependent on units a_j which are printed, as given in Table 7. The selling price of one issue of the journal is € 1.60. How many units should be printed if you have a choice to print 1, 2, ..., or 5 thousand copies?

Demand d_i	1,000	2,000	3,000	4,000	5,000
Probabilities p_i	0,40	0,30	0,20	0,06	0,04
Copies a_j	1,000	2,000	3,000	4,000	5,000
Cost $C(a_j)$	2,000	2,200	2,400	2,600	2,800

Table 7. Probabilities for demand for the journal and cost of copies

Calculation of the Expected Profit of a Single Decision. The cost is determined by the decision but the income remains subject to randomness. We might calculate its expected value in the sense of (4) in order to judge the present decision. We get a monetary value but cannot interpret it in isolation; we will have to calculate the expected profit also for the other possible actions and then compare the values.

Decision a_2: 2,000 copies with cost $C(a_j) = 2{,}200$			
Demand d_i	Probabilities p_i	Gross income	Net profit
1,000	0.40	1,600	–600
2,000	0.30	3,200	1,000
3,000	0.20	3,200	1,000
4,000	0.06	3,200	1,000
5,000	0.04	3,200	1,000
Expected profit (a_j)			360
Maximum loss (a_j)			–600

Table 8. Calculation of the consequences of a specific decision

The first entry of the *net profit* in Table 8 is –600 as when only 1,000 copies are sold, then the net profit is $1{,}000 \cdot 1.6 - 2{,}200 = -600$. It is important to note that an action cannot be judged in isolation: how could we interpret the expected profit of 360, or a maximum of loss of –600? A rational judgement requires the comparison of alternatives. Thus, we repeat the analysis for alternative numbers of copies (1,000 to 5,000) and arrange the results of computations in a combined matrix of net profits below (Table 9).

Comparing Different Decisions. We compare our options with respect to the number of copies according to the criterion of expected profit (rather than money we could also use utility of money).

Demand d_i	Net profit of decision: Number of copies a_j					Probabilities p_i
	1,000	2,000	3,000	4,000	5,000	
1,000	–400	–600	–800	–1,000	–1,200	0.40
2,000	–400	1,000	800	600	400	0.30
3,000	–400	1,000	2,400	2,200	2,000	0.20
4,000	–400	1,000	2,400	3,800	3,600	0.06
5,000	–400	1,000	2,400	3,800	5,200	0.04
Cost $C(a_j)$	2,000	2,200	2,400	2,600	2,800	Price per issue 1.60
Expected profit (a_j)	–400	360	640	600	464	
Maximum loss (a_j)	–400	–600	–800	–1,000	–1,200	

Table 9. Matrix of net profit depending on the decisions and the actual demand

The second column of Table 9 contains the random variable *profit* for the option of 2,000 copies, which was already displayed in Table 8. Other entries are calculated analogously. The probabilities for the entries in the second row are derived from those for the demand (they are contained in the outmost right column). According to the options we compare five different distributions for the profit; we could write them in separate tables and draw their bar graphs. As the probabilities are the same, we write the distributions compactly in the profit matrix.

It is obvious that no one would be willing to decide for option 1,000 as – whatever the demand will be – it will lead to a loss of –400. This reminds us to a principle of avoiding a sure loss (which is a basic principle in SJT). The option 2,000 delivers a positive expected profit of 360; however, it can lead also to an even higher loss of –600 as compared with the decision for 1,000 copies. This reflects a basic property of improving decisions under risk. Rarely can one find decisions, which are better throughout (whatever will happen), but it is easy to rule out inferior decisions. However, the remaining (admissible) actions cannot be compared to each other without a further criterion and what is the better decision depends on the criterion used. To improve a situation in one respect (to have a higher expected net profit) is accompanied by the risk of higher potential losses. One may even speak of an invariant in human life as seen from a general philosophical perspective on risk. The option 3,000, which yields an expected profit of 640, is better. However, it bears the risk of a loss of –800 (if demand is only 1,000, which has a probability of 0.40). It turns out that option 3,000 yields the maximum expected profit (640) and is – in the present model – the best decision.

The decision depends on the criterion used. If the criterion for the decision would be to minimize the maximum possible loss, then option 1,000 copies would be the best but this is not a feasible option at all. A minimax principle (minimizing the maximal possible loss) thus may lead to nonsensical decisions. Again, a general comparison is that fearing the maximum loss will often end up with a poor decision. If probabilities are modelled for the demand for copies of the journal (e.g., from past relative frequencies), then an expected profit criterion may be applied. Generally, this leads to better solutions but bears the risk of somewhat higher losses. In this way the Bayesian view would use risk in the sense of (4). The distribution of the demand is usually the result of validating a prior distribution on the demand by empirical data via the Bayesian formula and has a SJT character. Instead of basing the calculations on money, we could also attribute a utility to the different amounts of money.

Example 3. Medical Diagnosis

Medical diagnosis is a natural context to introduce conditional probabilities and reflect on the possible errors of making decisions. Suppose we have data from a group of 1,000 persons (see Table 10) about the cross-classification of their joint status of a disease *D* and the result of a biometric variable that is used to classify them as positive (T^+) or negative (T^-). The question is whether the result of the diagnosing variable can be used for diagnosing the disease or not.

	Status of disease		
Diagnosis	*D*	Not-*D*	All
T^+	9	99	108
T^-	1	891	892
All	10	990	1,000

Table 10. Data from an investigation in the disease *D* and a diagnosing method to detect *D*.

We imagine that each of the persons corresponds now to a ball that has two markers, one for the status of the disease and the other for the result of the diagnosis. We put all the balls into an urn and draw from it randomly. We get the following probabilities, which are easily read off Table 10:

a. $P(D) = \dfrac{10}{1000} = 0.0100;\ P(T^+) = \dfrac{108}{1000} = 0.1080,\ P(D \cap T^+) = \dfrac{9}{1000} = 0.0090.$

b. $P(T^+ \mid D) = \dfrac{9}{10} = 0.9000$ as compared to $P\,(T^+ \mid Not-D) = \dfrac{99}{990} = 0.1000.$

c. $P(D \mid T^+) = \dfrac{9}{108} = 0.0833$ as compared to $P(D \mid T^-) = \dfrac{1}{892} = 0.0011.$

We will use the name *prevalence* (incidence) for the probability of the disease. We can speak of the probability of a positive diagnosis to be 0.1080 in a. In b., we compare the probability of a positive diagnosis within "subgroup" D (0.9000; first column) to the analogue probability in "subgroup" Not-D (0.1000; second column) and state that the likelihood to get a positive result is 9 times larger in the group who have the disease D than in the group of people who do not have the disease. We can also give a name to the probability of a positive result T^+ in group D: this is the *sensitivity* of the diagnosing method. Its complementary probability is linked to an error of the diagnosis; here we have an error rate of 0.1000; i.e., on average we will not detect 10% of those who have the disease (as they get a negative result T^-). On the other hand, we can state that the diagnosis T^+ is wrongly attributed to a person without the disease with probability 0.1000. The complimentary probability is called the *specifity* of the diagnosis and it reflects the "reliability" that a person without D gets a correct negative result.

The context illustrates that diagnosing for the disease (or for the absence of the disease) is linked to risk. We can make two errors in deciding that a positive person has the disease and a negative person does not have the disease D. We can wrongly classify a person as to have the disease (first row, second entry) and we can wrongly classify a person not to have the disease (second row, first entry). The size of the errors, i.e., the probability of the errors characterizes the quality of the diagnosing procedure and whether the underlying biometric variable should be used for diagnosing D. Usually, the prevalence is estimated and the sensitivity and specificity are checked by a laboratory study of cases which are classified by other methods so that we certainly know of these persons their status of D. If we re-formulate the diagnosing problem as a statistical test of the null hypothesis "to have not D" and the alternative hypothesis "to have D", then the false positives (1 *minus* specifity) becomes the error of type I while (1 *minus* sensitivity) becomes the error of type II.

Elementary Approaches to Probability and Risk

Probabilities, especially conditional probabilities are difficult to perceive and hard to estimate and interpret in a specific context. Also, such probabilities reflect that we explore a real situation only via models that need not fit to it. There are interesting proposals to simplify the probabilistic data and to visualize them so that the information is more readable.

Gigerenzer's Tables with Natural Numbers

Gigerenzer (2002) has suggested the use of absolute numbers instead of (conditional) probabilities. All analyses are then based on these "natural frequencies" and are simplified to divide two numbers (in the sense of APT; see also Table 10). His research team has also investigated, which visualization is better to obtain and understand the solution, which is usually derived from calculations that are rarely understood (and are based on the Bayesian formula).

Suppose the data for the screening scheme for the detection of breast cancer are the following (Table 11; Gigerenzer, 2002): Of women between 40 and 50, 0.8% conceive cancer of the mammae yearly; we call this prevalence (incidence) of the disease. From those who have breast cancer, 87.5% are detected by anomalies in the mammogram (this is the sensitivity of the diagnosing procedure), i.e., diagnosed positive; from those who do not have breast cancer, 7% are also diagnosed positive indicating falsely that they have cancer (the specifity is 93%). In a probabilistic framework, a woman is randomly selected from the population and undergoes the mammogram. After she has a positive result, the question is what is the conditional probability to have cancer of the breast. An analogue question can be put after the mammogram is negative: what is this person's probability to have no cancer of the breast given that the mammogram is negative? The answer can be obtained by the Bayesian formula, which is quite formal. Gigerenzer suggests transforming all (conditional) probabilities to natural frequencies (expected values) for a group of 1,000 or more persons.

	Status		
	D	*Not-D*	All
T^+			
T^-			
All	8		1,000

	Status		
	D	*Not-D*	All
T^+	7	70	77
T^-	1	922	923
All	8	992	1,000

Prevalence (Prior): $P(D) = 0.008 = 0.8\ \%$;

Sensitivity: $P(T^+ \mid D) = 0.875 = 87.5\ \%$;

Specifity: $P(T^- \mid not\text{-}D) = 0.93 = 93\ \%$.

Table 11. Transforming (conditional) probabilities to natural frequencies by expected numbers in a statistical village

We round off the value of 922.56 to 922 for simplicity. The other data can just be filled by the usual side constraints. From the table of natural frequencies we can calculate any probability related to the diagnosis. If we restrict the selection to the first row, we read $P(D \mid T^+) = \dfrac{7}{77} = 9.09\ \%$, if we restrict the selection to the second row, we obtain $P(not\text{-}D \mid T^-) = \dfrac{922}{923} = 99.89\ \%$. Remarkably, the positive diagnosis has an unexpectedly low conditional probability for the disease, because of its low prevalence. The advantage of Gigerenzer's approach is that the formalism of Bayes' formula is circumvented and the final result is more convincing as only 7 from 77 positive diagnoses relate to women with breast cancer. In fact, Gigerenzer suggests that the expected numbers are arranged in a tree diagram instead of two-way tables as was done here.

Spiegelhalter's Icon Arrays

Spiegelhalter (2014) uses natural frequencies and presents each of the persons (units) in the statistical village by an icon. In his icon array he systematically compares the expected numbers of two scenarios, one with the screening scheme applied and the other without over a longer period (see Figure 1). This allows judging the benefit of the screening as compared to a decision for not undergoing the screening regime. Data in Table 12 display the expected numbers and illustrate the effect of PSA screening and digital rectal examination; numbers are for men aged 50 years or older, not participating vs. participating in screening for 10 years (1,000 men in each group).

Category of outcome	1,000 men without screening	1,000 men with screening	Icon used
Men dying from prostate cancer	8	8	●
Men dying from other causes	192	192	⊘
Men diagnosed for prostate cancer and treated unnecessarily	–	20	◉
Men without cancer that got a false alarm and a biopsy	–	180	◍
Men that are unharmed and alive	800	600	○
All	1,000	1,000	

Table 12. Expected numbers of diverse outcomes in a ten year screening scheme applied to 1,000 men as compared to the outcome of no screening (exact figures are slightly smoothed here)

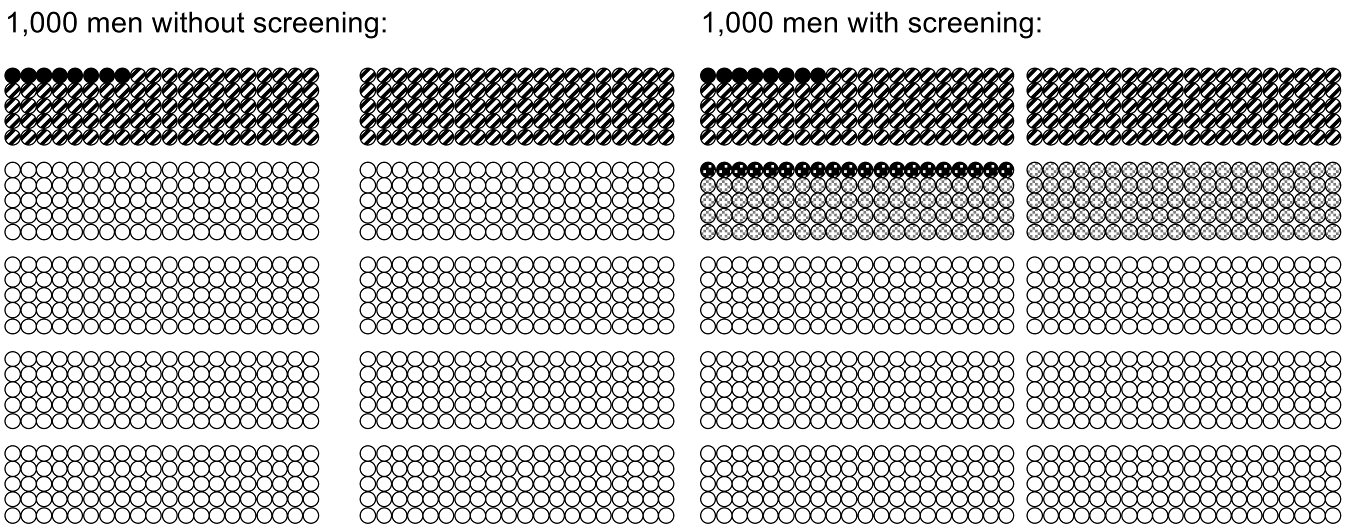


Figure 1. Icon array to illustrate the expected outcome for the prostate screening system as compared to no screening (from Spiegelhalter, 2014)

In the usual format one has to learn to read the information (and the formalism). Many formats in scientific communication hide more than they clarify the inherent information. Thus, there is an urgent need for elementary approaches. However, there are also disadvantages: The basic assumption in the approach of natural numbers is that people are all alike. On the one side there is a psychological barrier to accept such an averaging view. On the other side, there are not 1,000 men who are like me (as is suggested in the icon array). The data and the derived results when enhanced by the visual representation appear as factual and true. The key point is that whole numbers or visual representations of them can appear to give more validity than is appropriate as models and imprecise information are used. Both the probability of prostate cancer (the prevalence) and the sensitivity and specifity of the diagnosing procedures used in the screening regime are crude estimates for the underlying probabilities.

Odds and Probabilities

We can use odds for calibrating (tuning) our personal probability in the sense of SJT. Odds are the proportion of the probabilities that "*E* occurs to *E* fails to occur", i.e., $P(E):P(\overline{E})$. Odds for a six on a die are 1 : 5 and for a head on a coin, the odds are 1 : 1 (fifty-fifty). From odds, the probability may be recalculated by $P(E)=\frac{p}{p+q}$ so that $P(six)=\frac{1}{1+5}=\frac{1}{6}$.

In the diagnosing example from above we have prior odds for the disease of 8 : 992 or roughly 1 : 120. The probability of a positive diagnosis T^+ is 7/8 in the subgroup with *D*, and is only 70/992 in the subgroup without the disease *D* whence we state that the data likelihood for a positive diagnosis T^+ equals 7/8 : 70/992, or roughly 12 to 1; this expresses the fact that a positive result has a probability that is 12 times higher in the subgroup of women with breast cancer (*D*) than in the breast cancer-free group ($\overline{D}$). The formalism of the Bayes' formula can be reduced to multiply prior odds and data likelihood (see Borovcnik, 2012).

We represent the odds by fractions and multiply as if the factors are scaling factors for the odds (in the same way as scaling factors are multiplied in copying a sheet of paper repeatedly, scaling the intermediate copy up or down, to obtain the final scaling factor): $\frac{1}{120}\times\frac{12}{1}=\frac{1}{10}$, which yields posterior odds for *D* given the positive result T^+. The simplified posterior odds of 1 : 10 yield a posterior probability for *D* of 1/11, or 0.0909. The advantage of this calculus with probabilities is that we can identify the low prior odds of the prevalence to cause the surprisingly low probability to have breast cancer after a positive result. The diagnosing procedure per se is not very good but the data likelihood of T^+ increases prior odds by a factor of 12, which is not bad at all.

Evaluating and Sharing Risks between Different Stakeholders

Apart from personal risks, one may differentiate risks to the field where they arise (gambling, investments, etc.) or to those who are involved: single persons, or the society as a whole. In the latter case we might speak of societal risk. Climate change is a classic example of societal risk. Risks in medicine are harder to classify as there are single patients who are at risk but there is also an institutional side involved. The institutional side has its own goals which may differ from that of each individual.

Risks Shared between Stakeholders of Different Levels

For example, in public health there may be a discussion whether children should be vaccinated against measles or not. The health system is concerned about public order and is responsible to prevent an epidemic outbreak of the disease. It can build a frequentist scenario and weigh the different possibilities by probabilities and calculate the risks involved in wide-spread vaccination (possibly with a duty to partake) and reduced vaccination (no public financial support, no advertisement and counselling of parents). The individual parents have their subjective probabilities (own experience with the disease, strength of immune system, alternative ways to support health and cure in case of a disease, etc.). They cannot rely on the frequentist probability of the public health system, and the impact is hard to judge: the possible impact of the disease and the possibility of a bad consequence of the vaccination itself. While the parents stay with the personal consequences in either case, the doctor can be made liable for not informing about the possibility of the vaccination (or its negative side-effects). The pharma industries are a further stakeholder in this case, which has completely different financial interests and goals. We can see from this description that such decisions cannot be prescribed but have to be made personally, with all the assistance of public information that relates to the current state of medicine. For the different situations of the stakeholders in case of the recent HPV vaccination (that is intended to protect from cervical cancer) see Borovcnik and Kapadia (2011a).

Medical doctors enhance and increase their role by warning on health risks. With these warnings they are always on the safe side. If the patient follows their advice and the negative consequences do not occur, their concern seems justified; but if the patient still suffers from some consequences, there may be many other explanations and still the doctor's concern proves right. If the patient does not follow their advice and gets "ill", this seems again to be a confirmation of the doctor's concern; if the patient does not experience the negative outcome in this case, there will yet remain a reminder (you were only lucky!).

Societal Risks – Standards of Science and Society

In technological settings, the evaluation of risks has become a driving force to develop related concepts. While it seems desirable to keep technological risks as small as possible, economic reasons would be a "natural" limitation. And, there are no absolutely safe technologies so that the key question becomes whether we can reasonably decrease the risk.

When risk considerations are applied to complex systems (climate, ecosystems, world economy, etc.), the systems contain so many components so that their interactions become practically unpredictable. Whether climate change really establishes a risk remains doubtful for many people. The probabilities for the various developments are hard to assess and may be changed by time, technology, and economics. Many attempts are undertaken to simulate models (scenarios) on the basis of assumptions (with the well-known effects of flooding, increase of hurricanes, etc.) lacking serious attempts to estimate the probabilities of certain quantitative effects. The models resemble scenarios (what happens if …?) and cannot claim the status of a precise description of the real world as too many weakly supported hypotheses play a key role in their considerations.

This example reveals a key problem in presenting scientific results to the public. Standards that are viable within research may not be applicable for that generalization especially as the stakeholders have their own probabilities and are affected by the consequences completely differently. And there are dynamic developments involved that cannot be influenced further. For example, once there is a public decision for the energy use of atomic power, there are industries involved in the business and make their profit while on the other hand the public has to balance if an accident happens (as in Fukushima 2011, with the potential harm on health but also financially to cope with the damage).

The terms "objective" risk (when objective probabilities are involved) and "subjective" risk (when subjective probabilities are involved) are misleading, especially if the risk has to be shared between stakeholders of different levels. The calculated risks are only a support for finding a common decision but never can be used in the way that decisions are executed. Apart from the risk as it is per se, there is a perception of risk that is not based on (well-calibrated) subjective probabilities but more or less on intuitive short-cuts of such a risk calculation, which is based on beliefs and gut feelings.

When scientific results should be applied to other areas, especially to public decisions in society, we have to clarify the standards of evidence. In sciences, one essential strategy is to find hypotheses of specific relations between variables by rejecting a null hypothesis (that there are no such specific relations) and the type I error (of falsely rejecting the null) becomes vital. However, in applications a specific hypothesis has to be applied and the type II error of wrongly not rejecting the null hypothesis (of no special relations) becomes the nub of considerations. What if this type II error is very small? Then it might be very easy for an alternative hypothesis of special relations to get acknowledged by rejecting the null. We have dealt with these types of errors in a separate section earlier as they form part of the inner-mathematical approach towards risk within hypothesis testing.

Here an example illustrates matters. If a null hypothesis is rejected (because of a moderate or larger type I error), then one would establish a special relation between variables and use such a model for further action. What if this superimposed model were inadequate? Matters are complicated as nowadays it has become a standard to acknowledge empirical evidence as evidence for action. That means if data suggest a special relation, researchers usually leave out the next step of research, namely to find a substantial explanation from the context sciences for the special relation that has been prima facie found on statistical evidence. That would imply to build a functional (causal) model of interrelations between variables. One example here is the endeavour of the last two decades to find biomarkers that allow for early diagnosis of cancer of various types. Any association or correlation between such a biometric variable and the frequency of cancer has been taken seriously and investigated with millions of measurements and researches but – apart from Herceptin for breast-cancer (Breastcancer.org, n.d.) – with no apparent success. Medical researchers have not built a more refined model of the human metabolism (as in the Herceptin case) to explain the correlation (that has been found) on a context level.

It remains an open issue how to reconcile the standards of science and public implementation of scientific results. The institutional barrier between the stakeholders and the hugely differing interests are only two key elements that cause difficulties. There are further vital elements. There is the difference in information between stakeholders. The available frequentist (more objectivist) information might not be relevant for the single persons. Different criteria applied to the accessible information determine the decision. In the context of the journal copies we have illustrated the minimax principle (minimizing maximum loss) and optimizing expected values in the sense of (4). In general, these criteria lead to different decisions. Furthermore, there is neither universal agreement on such criteria nor a generally accepted scientific argument that one is superior over the others. The criteria fit different purposes, which can be related to tasks that involve risk.

Implications

Language is a highly sensitive issue for all activities in education and the problems are even more relevant in the case of probability and risk. The differing usage of risk has to be addressed in teaching explicitly. The diverse applications may help to enhance what is meant in each case and which connotation the ingredients have so that the results may be evaluated accordingly. The use of the term risk within statistical inference (both in the classical test and in the Bayesian framework) makes issues precise but has to struggle with two main problems:

i., the logic of statistical inference is complicated and misunderstood, which makes the terms more of a myth than precise.

ii., the use is different from everyday language and different from use in the context of economy and technical safety considerations. The overlap with everyday meaning and use of risk completes the description of the educational challenge.

To impose the technical jargon and meaning of risk to everyday use requires that the technical and mathematical concepts behind are fully perceived, which cannot be expected from the majority of students. Can changing the technical concepts be a viable alternative to prepare them better to everyday usage? Would we end up with a different statistics methodology, or new probabilistic and statistical concepts that are more suitable for the purpose of the individual? Or should we approach the problem from a more individualistic perspective from teaching and learning and yet aim to perceive the traditional concepts: increase the awareness of overlap in meaning, purpose, and naming the problems and concepts, and get alert to the wider use of similar wordings and notions in everyday contexts?

Psychological problems to separate between probability and utility have their consequences on proposals to introduce risk and utility considerations from early teaching onwards. Carefully chosen and deliberately discussed examples might help to clarify issues. Further psychological factors seem to be:

•The representation of a situation as win or loss influences decisions.

•The personal status of wealth seems to be vital but is largely unexplored.

•Rather than regarding the probability of the adverse events, people regard the impact of outcomes and possibly focus only on the worst case.

•Low probabilities are badly handled and provoke a shift in focus on impact (maximum loss) rather than evaluating a weighted risk in the sense of (4).

•Impact of various outcomes is reduced to impact in the worst case.

•Cumulative risk is often ignored though it may be relevant.

Constituents of risk are decisions; comparisons of decisions are easier to handle (and impose a partial ordering on various actions) than to understand a value attached to a single decision. How good an action is depends on the alternatives that are feasible. This fact can also be used manipulatively (by pre-set alternatives or reducing alternatives on purpose). We recapitulate some of our main themes.

a. The approach to probability (APT, FQT, SJT) to evaluate the outcomes (of which the adverse event is but one) regulates the character of calculated risks. However, seemingly more objective information might be useful for an institutional stakeholder but of little help for an individual.

b. The evaluation of the impact of the outcome may be in terms of money (win or loss), dichotomic (0 for correct decision and 1 for wrong decision), by utility of outcome, or even by idiosyncratic ways in judging specific unwanted events (as, e.g., in the three-door problem).

c. One-off decisions and repeated situations do require a different logic, which confuses and might provoke a focus on subsidiary "information".

d. Criteria used for decisions vary, e.g., for different stakeholders and are often very personal. Statistical information is abstract and often ignored while personal experience is vivid and tends to be overestimated; maximum loss seems more convincing than abstract weighting of impact by probabilities in the sense of (4).

The stakeholders' position has a definite impact on risk calculations and the decisions to make. There seems to be a clash for the logic of risk as implementation of considerations is different for the stakeholders; recommendations are also interest-driven rather than neutral in a scientific sense and differ also by diverse liabilities within systems (e.g., the health system, medical doctors, the individual patient, his or her relatives). What is rational and good for one stakeholder needs not be rational for another stakeholder, at least the connotation of the used concepts changes as does the information that can sensibly be used. Similar difficulties arise in risk-sharing in technological environments.

There are many topics that seem rewarding to put on the agenda, to ask back, and to find more time to clarify issues. Of course, it depends on the background of the learners whether such additional discussions are helpful for them. The many aspects of a concept may also confuse them. However, at least the teacher has to be aware of the circumstance that people have their private concepts (and which ones they have), that the use in everyday terms is ambiguous and diverse, and that even experts have widely diverging concepts and their own specific terminology. Not the least important element to note – experts also have their own interests and have to be challenged whether their advice can be taken as neutral. Statistics education is well-advised to be careful about the ways to teach issues of risk so that students can develop their own concepts and recognize the influential factors of their decisions as well as decisions that are made on a societal level for them.

Acknowledgements

We thank Ramesh Kapadia for valuable ideas and comments that helped to structure the present paper.

References

Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Chicago: Markham.

Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp.484-506). Barcelona: European Society for Research in Mathematics Education. Online: ermeweb.free.fr/CERME4/.

Borovcnik, M. (2011): Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (eds.), *Statistics in School Mathematics. Challenges for Teaching and Teacher Education* (pp.71-84). New York: Springer.

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didactica de la Matemática*, 2, 5-27. Online: www.aiem.es/index.php/aiem/.

Borovcnik, M., & Kapadia, R. (2011a). Risk in health: more information and more uncertainty. *Proceedings IASE Satellite Conference on "Statistics Education and Outreach"* (6pp.). Voorburg: ISI. Online: iase-web.org/Publications.php.

Borovcnik, M., & Kapadia, R. (2011b). Determinants of decision-making in risky situations. *Proceedings of the 58th World Statistics Congress* (6pp.). Voorburg: ISI. Online: 2011.isiproceedings.org/papers/950138.pdf.

Borovcnik, M., & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E.J. Chernoff, & B. Sriraman (eds.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education*, v. 7 (pp.7-34). Berlin: Springer.

Breastcancer.org (n.d.). How Herceptin works. Online: www.breastcancer.org/treatment/targeted_therapies/herceptin/how_works.

CCOHS (n.d.). Hazard and risk. *Canadian Centre for Occupational Health and Safety*. Online: www.ccohs.ca/oshanswers/hsprograms/hazard_risk.html.

Dubben, H.-H., & Beck-Bornholdt, H.-P. (2010). *Mit an Sicherheit grenzender Wahrscheinlichkeit. Logisches Denken und Zufall*. (With a probability coming close to certainty. Logic thinking and randomness.) Reinbek: Rowohlt.

Friedman, M. (1976/1962). *Price theory: a provisional text*. Chicago: Aldine.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2007). *Gut feelings: the intelligence of the unconscious*. New York: Viking.

Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.

Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Hansson, S. O. (2007). Risk. In E. N. Zalta (ed.), *Stanford Encyclopedia of Science*. Online: plato.stanford.edu/archives/spr2014/entries/risk/.

Kahneman, D., & Tversky, A, (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47 (2), 263-292.

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? *Journal of the American Statistical Association,* 88, (424), 1242-1249.

LeRoy, S. F., & Singell, Jr., L. D. (1987). Knight on risk and uncertainty. *Journal of Political Economy*, 95 (2), 394-406.

Merriam-Webster (n.d.). *Encyclopædia Britannica*. Online: http://www.merriam-webster.com/

Neumann, J. v., & Morgenstern, O. (1953). *The theory of games and economic behavior. 3rd ed.* Princeton: Princeton University Press.

Rasmussen, N., et al. (1975). Reactor safety study: an assessment of accident risks in U.S. commercial nuclear power plants, *WASH-1400*. Washington, DC: U.S. Nuclear Regulatory Commission.

Ritholtz, B. (n.d.). Defining risk versus uncertainty. *The big picture*. Online: www.ritholtz.com/blog/.

Sandblom, G., Varenhorst, E., Rosell, J. Löfman, O., & Carlsson, P. (2011). Randomised prostate cancer screening trial: 20year follow-up. *British Medical Journal*, 342. Online: doi:10.1136/bmj.d1539.

Spiegelhalter, D. (2014). What can education learn from real-world communication of risk and uncertainty? Invited lecture. *8th British Congress of Mathematics Education* (BCME 8), Univ. of Nottingham. See also: *Harding Center for Risk Literacy*. Online: https://www.harding-center.mpg.de/en/health-information/facts-boxes/psa.

Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, vol.4(1): Personelle Wahrscheinlichkeit und Rationale Entscheidung; (2): Personelle und statistische Wahrscheinlichkeit* (Problems and results of science and analytic philosophy). Berlin: Springer.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211 (1), 453-458.

A Cognitive Framework for Normative Reasoning under Uncertainty, and Reasoning about Risk, and Implications for Educational Practice

Sylvia Kuzmak

Rise Coaching and Consulting LLC NJ, USA

Abstract: Clarifying what is normative or appropriate reasoning under various circumstances provides a valuable reference for guiding what should be taught, and, in contrast, what should not be. This paper proposes a cognitive framework for viewing normative reasoning and behavior under uncertainty, including the applying of knowledge of probability and statistics in real world situations; and identifies implications for educational practice. Factors relevant to normative reasoning under uncertainty that are addressed within the framework include: risk of misapplying statistics knowledge, involvement of mathematical and non-mathematical reasoning, knowledge of real world domains and situation/application detail, and existence of expert consensus. The cognitive framework is illustrated using examples of reasoning about risk, including industry standards for risk management. The work of Kahneman and Tversky, G. Gigerenzer, and others is related to and contrasted to the framework presented.

Keywords: reasoning under uncertainty, statistical reasoning, probabilistic reasoning, risk, improving probability education, improving statistics education, misapplication of statistics, statistics application.

Introduction

Clarifying what is normative or appropriate reasoning under various circumstances provides a valuable reference for guiding what should be taught, and, in contrast, what should not be. This paper proposes a cognitive framework for viewing normative reasoning and behavior under uncertainty, including the applying of knowledge of probability and statistics in real world situations; and identifies implications for educational practice. In sections below, factors relevant to normative reasoning under uncertainty are identified, illustrated with examples, and related to the research literature on reasoning under uncertainty. In particular, examples involving reasoning about risk are addressed, including reasoning reflected in industry standards for risk management. In the final sections, the factors are integrated into a cognitive framework; and implications for educational practice are identified.

In real world situations, we are often in the position that the outcome of a situation, which is subject to uncertainty, matters to us. To reason and behave normatively at such times is important, since doing so helps to bring on potential benefits and/or to stave off potential difficulties. Such real world situations draw interest and are engaging, and typically call for action, because the results matter. Such real world situations also are aptly described as involving risk. Not only is the outcome of the situation uncertain, with alternative possible outcomes, but the possible outcomes have positive or negative impact, so that there is risk that a positive outcome will not occur, and/or risk that a negative outcome will occur. By addressing what is normative reasoning and behavior in such situations, the cognitive framework presented here applies in general to reasoning about risk.

As a simple example, consider observing the rolling of a pair of six-sided dice. There is uncertainty in the outcomes, but unless the rolling occurs in the context of a game or other real world consequences, the outcomes don't really matter. Now, consider that the rolling of the dice is occurring in the context of gambling, and that you are about to place a large sum of your money as a bet on the

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 140-156

outcome of the roll. Now the outcome is more important, the situation is more engaging, you are interested in your options for action that may make a difference in the situation, and there is risk. As another example, consider that you are a young person and occasionally contemplate your own mortality, but realize that, due to your generally safe environment and healthy habits, your odds are good that you will live a long life; and so the issue of your possible early death is not of real concern to you. Now consider that you and your spouse are just starting a family. Although the probability that you will die relatively young is still low, now the possibility of your early death is of concern to you, since it would have a great financial and otherwise life-impacting effect on your remaining family. The situation regarding your mortality is now more important, more engaging, involves risk, and has led you to consider possible actions, including buying life insurance.

The focus of the framework presented here is not just on people's judgments of probability of outcomes in situations involving uncertainty, but a broader sense of reasoning under uncertainty that includes consideration of risk, perceived consequences of outcomes, and human actions/ behavior in that context.

Mathematical and Non-mathematical Reasoning

Historically, the research area of "reasoning under uncertainty" in cognitive psychology and decision science has been closely identified with the mathematics of probability and statistics. In a seminal paper in the field, Tversky and Kahneman (1974) reported their research in which adults had been posed written problems calling for them to reason and make judgments under uncertainty; and the authors concluded that their subjects showed "biases" and "errors" in their judgments, using reasoning heuristics, such as *representativeness* and *availability*, while not being influenced by relevant mathematical information provided in the problems, such as prior probabilities or base rates, and sample sizes. The authors describe the representativeness heuristic as judging the probability that an object A belongs to a class B, or that an event A originates from a process B, "by the degree to which A is representative of B, that is, by the degree to which A resembles B" (p.1124). An example provided is the case of being given a written personality sketch of a person and being asked to judge the probability that the person has a particular occupation, such as librarian, engineer, or lawyer; and making the judgment based on the person's similarity to one's stereotype for the occupation. The availability heuristic refers to assessing "the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind" (p. 1127). An example provided is the case of "assess[ing] the risk of heart attack among middle-aged people by recalling such occurrences among one's acquaintances" (p.1127). The body of research on heuristics and biases in reasoning and judgment has grown over the decades (for example, see compilations of research in Kahneman, Slovic, & Tversky, 1982; and Gilovich, Griffin, & Kahneman, 2002). A continuing theme within the body of research has been to note the presence of error in human judgment under uncertainty with respect to not understanding basic principles of probability and statistics and/or not applying them when appropriate. In a review of his work and others' on heuristics and biases, Kahneman (2011) noted that a takeaway of the work is "that there are distinctive patterns in the errors people make" in their "judgments and choices" (p.3).

Others have objected to the negative view of human ability and performance that has come along with the "heuristics and biases" literature, and have observed that people, in their daily activities do regularly encounter uncertainties, and show competence in dealing with them, although they may use heuristic reasoning, and not be applying a knowledge of probability and statistics in these situations. For example, Cohen (1981) noted regarding research on heuristics and biases, that subjects' performance on unfamiliar tasks and in laboratory conditions, is not an adequate basis for judging their rationality or competence; and that, in some cases, subjects may legitimately interpret the problem differently than does the experimenter, and respond rationally within their interpretation. Also, he noted that some of the research is merely a "test of intelligence or education" (p.325), demonstrating "a lack of mathematical or scientific expertise" (p. 325) that is unreasonable to expect that all would have. Kahneman (2011) has defended the negative focus of the heuristics and biases literature by drawing an analogy to the legitimate attention to the study of disease in the field of medicine (p.4).

To call out "error" in judgments and choices, as is done in the heuristics and biases literature, is to presume some notion of what is normative or appropriate to do under the circumstances in which the error was observed. Cohen (1981) criticized researchers' tendency to assume an "inappropriate normative theory" (p.328) drawn from logic and statistics, not recognizing that many important normative issues are still controversial among experts in the field. Lopes and Oden (1991) criticized the heuristics and biases literature for its "normative models drawn from probability theory, economics, and logic" (p.201), inadequate to serve for the broad range of human decision making under uncertainty. They noted that, in the field of artificial intelligence, where intelligent performance of machines is the goal, heuristic reasoning is viewed positively and researchers "seek … out and embrace" (p.209) the use of heuristics. Lopes and Oden advocate the positive perspective that observed human performance "signal[s] the operation of a quite different kind of intelligence than is implied by conventional notions of rationality, an intelligence reflecting the properties of the [human's] massively parallel computational system that has evolved to meet the requirements of existence in a noisy, uncertain, and unstable world" (p.201).

Gigerenzer (1996) has critiqued the heuristics and biases literature as having too narrow a view of normative reasoning, and being too vague regarding the cognitive processes involved in heuristic reasoning. Regarding the narrow view of norms, he noted that, within the literature, there is a "practice of imposing a statistical principle as a norm without examining content" (p.593) of the problem posed; he says, "A convenient statistical principle, such as the conjunction rule or Bayes's rule, is chosen as normative, and some real-world content is filled in afterward, on the assumption that only structure matters. The content of the problem is not analyzed in building the normative model, nor are the specific assumptions people make about the situation" (p.592). So the failure of researchers to more broadly analyze the content of the problem and reasonable potential interpretations, then bypasses analysis of information that is needed to determine what is normative reasoning under the circumstances. As an example, Gigerenzer noted that the words "probable" and "and" used in a written problem, may legitimately be interpreted with natural language meanings by subjects, while researchers intend their interpretation as mathematical or logical terms; and the subjects' reasonable interpretations bear on what is normative reasoning under the circumstances.

Instead of viewing heuristic reasoning as being subject to error due to ignoring some information, Gigerenzer and Gaissmaier (2011) make the point that the use of heuristic reasoning that "ignor[es] part of the information can lead to more accurate judgments than weighting and adding all information, for instance for low predictability and small samples" (p.451). They note that "for many decisions, the assumptions of rational models [defined by logic or statistical models] are not met, and it is an empirical rather than an a priori issue how well cognitive heuristics function in an uncertain world" (p.451).

Regarding normative reasoning under uncertainty, then, a conclusion here is that both *mathematical reasoning* (using mathematical logic, probability, and statistics) and *non-mathematical reasoning* (such as heuristic reasoning and language interpretation) have a role. Non-mathematical reasoning is used here as a broad umbrella term to contrast to mathematical reasoning, the latter referring to reasoning using the formal subject matter of mathematics, including concepts, principles, and techniques of logic, probability, and statistics. It should be noted that what is called here non-mathematical reasoning nonetheless may have quantitative aspects, such as judging degrees of similarity and frequency. It is beyond the scope of this paper to provide an inventory and elaborated description of this rich and important category that is simply bundled here as non-mathematical reasoning. For many real world problems involving uncertainty, probability and statistics are powerful tools applicable to the circumstances, and are widely acknowledged as the norm to use under those circumstances. However, there are also situations involving uncertainty in which models based on probability or statistics do not apply, e.g., when the required assumptions for the model are not met; and yet, effective reasoning leading to adaptive judgments and behavior may proceed using heuristic reasoning. For some situations, there may look like a way to apply mathematics, but that application of mathematics in the situation may not be normative. In the later section that introduces the cognitive framework for normative reasoning and behavior under uncertainty, an example is presented which illustrates the need for caution in applying mathematical reasoning and the role of non-mathematical reasoning in normative reasoning and behavior.

Risk of Misapplying Knowledge of Probability and Statistics

The power and value of knowledge of probability and statistics lies in its appropriate application in real world situations. If the applicability of the math in certain circumstances is controversial, or if one applies the math to a situation without care to attend to relevant detail, or if one's knowledge is partial or lacks the firmness to ensure appropriate application to the circumstances, then there is real potential for misapplication. In such cases, to apply such knowledge involves risk and calls for awareness of the risk.

Misapplications of knowledge in probability and statistics may also occur intentionally with self-serving motivation, what has been called "lying with statistics," done by those who aim to take advantage of an audience who is inattentive, unsuspecting, and/or weak in statistics knowledge, with the goal to reap advantage for themselves to the detriment of others. Due to this possibility, the use of reported statistical information in reasoning involves risk and calls for awareness of the risk.

The risk of misapplication is greater, the greater the likelihood that the application of knowledge is in error, and the greater the negative impact of the error(s). Regarding negative impacts, misapplications may lead to flawed statistical results, e.g., inaccurate estimates or false reports of statistical significance of results. And then those flawed results may be used as a factor in decision-making, leading to unsound decisions with follow-on negative consequences. Whether the situation relates to results in medical science and patient treatment choices, business data and product promotion decisions, political polling and decisions on allocation of campaign resources, environmental impact data and setting of government environmental policy, product information in consumer reports and personal purchasing decisions, or other matters, the negative consequences of misapplication of knowledge of probability and statistics may be great.

Regarding the likelihood of a person's application of probability and statistics in a situation being in error, relevant is the chronic finding concerning modern education in probability and statistics that the subject matters are difficult for students to learn and for teachers to effectively teach. In a frequently cited research review from over 25 years ago, Garfield and Ahlgren (1988) concluded: "… despite the enthusiastic development of new instructional materials, little seems to be known about how to teach probability and statistics effectively" (p. 45). In a more recent research review, Tishkovskaya and Lancaster (2012) conclude that "Despite the widespread emphasis on reform in the teaching of statistics and the increase in papers on statistics education in the research literature, statistics is still viewed as a discipline with a need for significant improvement in how students are educated (Garfield & Ben-Zvi, 2008)" (p. 2).

Considering the circumstances just reviewed, including widespread weak knowledge of probability and statistics, yielding potential for misapplication of probability and statistics, and the range of negative impacts that may result from misapplication; therefore, in present circumstances, it is reasonable to view the risk of misapplying probability and statistics as significant.

There is evidence that even those who have taken courses in probability and statistics as part of their professional training, and who use statistical methods in their profession, fail to apply those methods appropriately, to the detriment of their professional work (Simmons, Nelson, & Simonsohn, 2011; Ioannidis, 2005; John, Loewenstein, & Prelec, 2012). This is presumably due at least in part to problems in learning the subject, and may also occur intentionally with self-serving motivation, e.g., manipulating data to be able to report statistically significant results and get one's work published. It is reasonable to suppose that the follow-on negative consequences of such professional misapplication of probability and statistics, including wasted effort and opportunity lost, incurred across the broad range of professional fields, is great.

In this paper, I raise the point that it is right to take into account in reasoning under uncertainty the risk of misapplying knowledge of probability and statistics. In a sense, the mathematics is a double-edged sword—powerful when applicable and applied appropriately, and also presenting great risk, both from unintentional misapplication, and from intentional misuse. Regarding the unintentional misapplication, it is a virtue to appreciate the limits of one's own state of knowledge. Incompleteness and lack of firmness in one's own knowledge is a true source of uncertainty, as one analyzes a situation and faces decision-making under uncertainty. Self-awareness and accuracy of perception of one's own weaknesses in knowledge supports normative reasoning and behavior for the individual in those circumstances. Recognizing the virtue of such awareness is akin to appreciating the wise saying, "A little knowledge is a dangerous thing." Regarding the intentional misuse, given the reality of its presence (e.g., in commercial advertising, political campaigning, and business financial reporting), again, it is a virtue to consider this factor in reasoning under uncertainty, being circumspect in evaluating reported statistical results and in integrating them into one's own reasoning.

Illustrations and Discussion

To illustrate the potential for error in applying knowledge of probability and statistics, two real world examples are described below, that involve drawing inferences about a population from a sample. Discussion follows regarding the potential errors and negative consequences.

Two examples. Consider the case of conducting an opinion survey and compiling the results; or sampling products from a manufacturing process, and measuring the sample's characteristics. In the case of the survey, for each item on the survey, the percentage of people in the sample making each possible response is tabulated. In the case of the manufacturing process, for each sampled product, measurements are taken for a set of characteristics of the product, and the mean measurement (arithmetic average) for each of the characteristics (across the products in the sample) is calculated. One may now conclude that the opinion percentages in the sample tell us the opinion percentages in the population as a whole (or close to it); and one may conclude that the manufacturing process, in general, produces products with the mean measurements obtained from the sample (or close to it); but if one were to do so, one would be misapplying knowledge of probability and statistics. What are the errors?

Need evidence that the sample is representative of the population. Steps need to have been taken to ensure that the sample is representative of the population about which one wishes to make inferences. In the case of a survey, this means that the demographic characteristics of the sample should match the demographic characteristics of the population. So, for example, one should sample the full age spectrum, and not just residents at a retirement community or just newly registered voters. If one wants to predict the outcome of an election, then one should sample likely voters (ones that have a history of voting regularly), and certainly not just unlikely voters. In the case of the manufacturing product sample, one should sample from the range of products about which one wants to draw conclusions, and not just the products that are least costly to sample, or for which one expects the "best performance" for the characteristics; and one should sample from the full spectrum of manufacturing conditions, e.g., from both day and night shifts, and from a range of typical manufacturing equipment conditions. To summarize, the procedures for data collection and analysis should be designed and be faithfully executed to ensure that the sample is representative of the population. In reality, samples may be collected in an unplanned manner; or in a manner to minimize collection time and/or cost, without regard to the resultant representativeness of the sample; or in a manner designed to manipulate results to a desired end instead of focusing on having a representative sample. Large data sets may be collected using multiple personnel, each using different and undocumented procedures; and respondents in a study may be anonymous, without demographic information collected; and such circumstances limit the usefulness of the data set for making inferences about the population. Random sampling from a population is a technique to obtain a representative sample, but that technique has not necessarily been faithfully used.

The mean can be a misleading measure of central tendency. Just knowing the mean (arithmetic average) of a sample, or of the entire population of focus for that matter, does not tell one what the underlying distribution of the population is. Although one may expect the bulk of a distribution to be around the population's mean, and that may often be true, such as for a normal distribution; it is also possible for it not to be true, such as is the case for bimodal or highly skewed distributions. For example, household income in the United States is a highly skewed distribution, with a long tail extending up into the super rich range; and the population mean income is much higher than the income of the central bulk of the population. For that reason, the median and not the mean is used in government reporting of US household income trends. In the case of an opinion survey, one may have a bimodal distribution if two groups comprising the population respond very differently to a question (e.g., teenagers and adults). In the case of measuring products from a manufacturing process, one may have a bimodal distribution if settings of the manufacturing machinery shift abruptly (e.g., accidentally due to human error, or due to mechanical failure) during the time when the sample is being collected. In these examples, the mean of the sample is a misleading measure of central tendency, since it does not inform one about where the bulk of the population lies, and, indeed, it may be in a range where very little of the population falls.

Need to take into account the sample size, whether sampling is random, and population variance. Since there is variation in a population, evident in its distribution, a single observation taken randomly from the population may be from anywhere in the possible range. When a random sample is taken from any population, the larger the sample size, the sample mean approaches being normally distributed with the mean being the same as for the population from which the sample is drawn (by the classical Central Limit Theorem). Thus, in the examples, for the sample mean to be informative regarding the population mean, the sample must be random and of sufficient size; and the larger the variance of the population, the larger the sample size is needed to bound the estimate of the population mean. The general caution is not to make wide-sweeping conclusions about a population based on a few observations or a small sample, and when the sample is not random.

Need to take into account required assumptions when applying statistical techniques. Based on an opinion survey, one might want to assess whether teenagers and adults differ in their opinions on a particular issue. Or, similarly, based on manufacturing product measurements, one might want to assess whether production quality differs for the day shift and the night shift. Both these inquiries relate to whether two groups are different. There are statistical tests designed to assess whether two groups are different, and these tests have assumptions that apply for their use. A commonly used test for difference between two groups is the Student's *t*-test, which assumes the data from the two populations are each normally distributed with the same variance. If the assumptions for the use of a statistical test do not hold for the situation of study, then to use that statistical test in that situation is not appropriate, and to do so may yield misleading results. Data is not always normally distributed, so use of techniques that assume normally distributed data is not always appropriate. For testing whether two groups differ, there are non-parametric statistical tests that apply no matter what the two population distributions are, such as the Mann-Whitney *U*-test, which is based on ranking of scores. Unless the assumptions for a technique's use are met in the situation to which it is applied, the results may be misleading, and lead to unsound decisions and negative follow-on consequences.

Need to consider practical significance along with statistical significance, notably when sample size is large. Continuing the example of looking for whether there are differences between two groups, and assuming that an appropriate statistical test is applied, let us say that a statistically significant difference is reported. So, for example, for the opinion survey, teenagers and adults are found to differ statistically significantly in their support for increased federal funding for low interest student loans. For the manufacturing product measurements, product quality is found to be statistically significantly different for the day and night shifts. A natural conclusion in both these cases may be that the results are meaningful, relevant, and noteworthy. However, that may not be the case. It is not just the statistical significance of the difference that matters, but also the magnitude of the observed difference between groups. A very small difference between groups may be detected by a statistical test if the sample size is large enough. In the two examples, if the sample size is large (e.g., 200), the actual magnitude of the observed difference between the two groups (that are statistically significantly different) may be quite small, not large enough to warrant treating the two groups differently, or warranting any action based on the observed difference between groups. An example of an abuse related to this principle is when a researcher seeks to conduct a study that will yield statistically significant results so that it will be publishable, and designs a study for which there is low effort and cost to include additional subjects or items in the study, accumulates large sample sizes for the two groups, and obtains results that show a statistically significant difference between groups, and show only a small observed magnitude difference between the groups, thus producing results that are of limited usefulness or merit.

The above real world examples and discussion provide a brief view into specific potential misapplications of knowledge of probability and statistics and potential follow-on negative impacts, to illustrate that the risk of misapplication of knowledge is broadly present and practically significant in real world situations. It is a true source of uncertainty, of which it is a virtue to be aware and to address with reason; and, it is right to be included in a description of normative reasoning under uncertainty.

Real World Domain Knowledge and Situation/Application Detail

Reasoning under uncertainty involving the application of knowledge of probability and statistics involves reasoning about the real world situation to which the math is applied. The illustrations and discussion in the previous section illustrate that mathematical concepts, principles, and techniques are mapped onto a real world situation, and reasoning proceeds within the domain of the situation, referencing relevant details of the situation and the mathematics that applies. For example, the concept of random selection must be applied within the application detail of selecting and procuring people to participate in an opinion survey, or within the detail of establishing a sampling plan of products in a manufacturing process. As another example, the population about which one wishes to draw inferences by sampling needs to be defined, in terms of details in the application domain, such as all voters in the state who will vote in the fall election or all voters in the county who are eligible to vote, or all products manufactured at a particular manufacturing site or product lines with high levels of consumer complaints.

Indeed, normative reasoning under uncertainty within a particular domain is dependent on the application details. For example, if one is going to conduct an opinion survey (or interpret the results of one) appropriately, part of the relevant knowledge is knowing factors within the domain that are relevant to ensuring a representative sample. If one does not apply appropriate domain knowledge in reasoning under uncertainty, then the knowledge of the math, out of the context of the application details, cannot deliver the normative reasoning leading to sound judgments for the situation.

In summary, domain knowledge and expertise are important contributors to normative reasoning under uncertainty within the domain. For example, important within medical science is knowledge and expertise in medical and pharmaceutical research, important within business is knowledge and expertise in market surveying and manufacturing, and important in opinion polling is knowledge and expertise in opinion research.

The close relationship between the math of probability and statistics, and domain knowledge in the areas to which it has been applied, is illustrated in the history of the development of the mathematics of probability and statistics. Gigerenzer et. al. (1989) have noted: "Perhaps more than any other part of mathematics, probability theory has had a relationship of intimacy bordering upon identity with its applications. Indeed, there was arguably no 'pure' theory of mathematical probability until 1933 …, and until the early nineteenth century, the failure of an application threatened the theory itself … For much of its history, probability theory *was* its applications" (pp. xiii-xiv). As probability theory spread to new domains, "…the mathematical tool shaped, but was also shaped by, its objects" (p. xiv). "From its beginnings in the mid-seventeenth century, probability theory spread in the eighteenth century from gambling problems to jurisprudence, data analysis, inductive inference, and insurance, and from there to sociology, to physics, to biology, and to psychology in the nineteenth, and on to agronomy, polling, medical testing, baseball, and innumerable other practical (and not so practical) matters in the twentieth" (p. xiii). The close relationship between the math of probability and statistics and the real world application domains to which it has been applied, is also evident in the cognitive realm: normative reasoning under uncertainty in real world situations incorporates and integrates both knowledge of the math of probability and statistics, and knowledge of the application domain.

Expert Consensus

As discussed in sections above, to apply the math of probability and statistics to a real world situation involving uncertainty, is not always the normative approach to reasoning for the situation. Sometimes the assumptions required to apply the math do not hold in the situation, so the math is not applicable; and sometimes, with either good or bad intentions, math knowledge is misapplied in a situation. Rather, what determines what is considered normative reasoning under uncertainty in various situations is the consensus of experts within the culture or relevant communities.

Among mathematicians in probability and statistics, there has not always been consensus. Gigerenzer et. al. (1989) have described the long and heated disagreement between giants of the field R. A. Fisher and the duo J. Neyman and E. Pearson regarding tests of significance and hypothesis testing, and note that the "vigorous controversies … have not ended" (p. 105). They added that "Disputes no less heated have characterized the relationship between Bayesians and frequentists" (p.105). Lopes (1982) reviewed literature related to the understanding of randomness, and raised the point that some assume "that randomness is clearly defined and well understood by those who are not naïve [mathematicians and philosophers]. Nothing could be farther from the truth" (p.628). She goes on to describe the different views of randomness of R. von Mises, G. Spencer-Brown, K. Popper, and others. Cohen (1981) made the point that the consensus of experts in a field can change over time, and cited the example of challenge arising to the Frege-Russell logic of quantification that seemed once "a universally received doctrine" (p.328).

As discussed in sections above, in reasoning about real world situations involving uncertainty, it is not just reasoning based on the math of probability and statistics that may apply, but necessarily also applicable is non-mathematical reasoning, such as heuristic reasoning, language interpretation, and the use of domain knowledge for the situation. Relevant experts in this case are both mathematical statisticians and experts in the domain of the application. Normative reasoning under uncertainty in the domain is determined by a consensus of these experts, if such a consensus is reached. If the math of probability and statistics is not clearly applicable within a domain of application, then normative reasoning under uncertainty in the domain using non-mathematical reasoning is determined by consensus of experts in the domain, again, if such a consensus is reached.

Normative Reasoning for Risk Management

Risk management involves reasoning under uncertainty and is an important function in business and industry. Industry standards for risk management have been established by industry-supported organizations, including the International Standards Organization (ISO) and, in the US, the federal government-sponsored Software Engineering Institute (SEI), both of which have assembled groups of experienced practitioners from industry (domain experts) to collaborate to establish standards for risk management, and to maintain those standards over the years. Established by domain experts and accepted by the community, the standards describe normative reasoning and behavior for risk management, and are used widely in industry as guidelines to promote effective practice of risk management.

ISO 31000:2009 Risk Management standards. The purpose of *ISO 31000:2009 Risk Management – Principles and guidelines*, as described on the ISO website, is to help "any organization regardless of its size, activity, or sector" to "increase likelihood of achieving objectives, improve the identification of opportunities and threats and effectively allocate and use resources for risk treatment." Within the standard, the risk management process includes risk assessment (identifying, analyzing, and evaluating risks) and risk treatment, all preceded by establishing the context for risk management. Risk analysis includes considering "the causes and sources of risks, their positive and negative consequences, and the likelihood that the consequences can occur" (p.21). A companion standard *ISO/IES 31010:2009 Risk Management – Risk assessment techniques*, "provides guidance on selection and application of systematic techniques for risk assessment" (Scope section, para. 1). To support identifying the risks that should be managed, the standard provides guidelines for selecting risk assessment techniques, and describes a set of 31 risk assessment techniques from which one may select, including brainstorming, structured or semi-structured interviews, Delphi Technique, check-lists, and Root Cause Analysis.

Regarding risk treatment, the ISO 31000:2009 standard provides a list of possible ways to manage or treat the risks one has identified:

- avoiding the risk by deciding not to start or continue with the activity that gives rise to the risk;
- taking or increasing risk in order to pursue an opportunity;
- removing the risk source;
- changing the likelihood;
- changing the consequences;
- sharing the risk with another party or parties (including contracts and risk financing); and
- retaining the risk by informed decision. (p.9)

The standard illustrates use of heuristic reasoning and the deep involvement of domain knowledge in the specification of the standard.

CMMI for Development, Risk Management standards. The SEI's Capability Maturity Model Integration (CMMI) for Development (Chrissis, Konrad, & Shrum, 2011) is a "collection of best practices that help organizations to improve their processes" (p.xv), originally established in its earliest form (CMM) for software development organizations in 1995, but also applicable to other organizations. CMMI-DEV includes a process area for Risk Management, with the purpose "to identify potential problems before they occur so that risk handling activities can be planned and invoked as needed across the life of the product or project to mitigate adverse impacts on achieving objectives" (p.481). The risk management practices include identifying and analyzing risks, including evaluating risk likelihood and risk consequence (i.e., impact and severity of risk occurrence) through human judgment, and tracking risks to monitor whether they have reached pre-planned thresholds to trigger management activities, including implementation of risk mitigation plans. The best practices illustrate the use of heuristic reasoning and involvement of domain knowledge, e.g., in the identification and quantification of risks and in the establishment of risk mitigation plans.

A Cognitive Framework for Normative Reasoning and Behavior under Uncertainty

In the previous sections, factors relevant to normative reasoning under uncertainty for real world situations, have been identified and discussed. In this section, the factors are integrated into a cognitive framework. The purpose of the framework is to present, in integrated form, key features of normative reasoning under uncertainty for real world situations, to further emphasize their role, and show how they tie together. The intended message is not that this is a complete list of features, but rather, a fundamental set of features that deserves emphasis, and provides useful perspective for researchers and educators. At the end of this section, the features and how they tie together are illustrated through an example.

Regarding the scope of the framework, it focuses on addressing human performance in real world situations involving uncertainty, in which what may happen or may be true, matters; and not with abstract, artificial, or simplistic tasks, e.g., as may occur with laboratory subjects being presented with short written problems to which to respond. In real world situations, the focus is on performing well or adaptively given the circumstances one is faced with and the consequences that may result. Real world situations are engaging, rich in detail, and call for action, and are where the value of normative reasoning under uncertainty delivers its payoff.

Real world situations may be categorized into "domains," e.g., involving medical science, business, politics, or consumer behavior. Often, within a domain, one can identify expert practitioners who may collaborate and reach consensus on what is normative reasoning under uncertainty for types of situations in the domain. Expert consensus provides the basis for identifying reasoning and behavior as normative. In contrast, to assume that any application of probability and statistics to a real world situation is normative, is unfounded. Also, to assume that every real world situation has an established normative standard for reasoning and behavior is unfounded.

For domains of application of probability and statistics, expert practitioners do exist, and provide a basis for establishing normative standards for reasoning under uncertainty within the domain, including normative application of probability and statistics. Still, there are controversies among experts, and so one cannot assume that there are normative standards for all applications of probability and statistics nor for all situations within a domain. What is and is not normative application of probability and statistics within a domain then has implications for what should and should not be taught, which is addressed in the last section of this paper on educational implications.

A Cognitive Framework for Normative Reasoning and Behavior under Uncertainty				
Steps:	**Get initial situation understanding**	**Identify and evaluate applicable reasoning threads**	**Apply synthesizing reasoning and resolve understanding**	**Respond/ act, given resolved understanding**
Actions (cognitive and behavioral):	• Get initial situation understanding including situation/ application detail • Recognize elements/ sources of uncertainty • Focus on question(s) about the situation	• Identify relevant information and reasoning threads to address the question(s) • Seek out additional information about the situation to clarify or confirm the situation aspects related to reasoning threads; for mathematical reasoning, confirm that required assumptions hold	• Apply synthesizing reasoning to reasoning threads, leading to resolving understanding • Seek out additional information about the situation to clarify or confirm the situation aspects related to synthesizing reasoning threads; for synthesizing using mathematical reasoning, confirm that required assumptions are satisfied	• Based on resolved understanding, respond/ act, e.g.: judge situation (e.g., make an estimate), choose an alternative, act to change the situation, continue to track situation to monitor uncertainties, abandon/avoid situation
Character-istics:	• Normative reasoning is agreed upon by expert consensus within a domain of application and accepted by the community • Reasoning incorporates and integrates: mathematical reasoning (based on logic, probability and statistics), non-mathematical reasoning (e.g., heuristic reasoning and language interpretation), knowledge of domain and situation/ application detail			
Performance:	• Proficient practitioners provide examples of normative reasoning and behavior • For individuals not proficient in normative reasoning for a situation (reasoning based on expert consensus), in general, it is a personal norm to consider the risk of misapplying knowledge, in one's reasoning and behavior for the situation			

Figure 1. A cognitive framework for normative reasoning and behavior under uncertainty

Having reviewed its intended scope, a cognitive framework for normative reasoning and behavior under uncertainty, expressing and integrating the factors discussed in previous sections, is presented in Figure 1 (above).

In the cognitive framework, the process for reasoning and behavior is broken down into four high level steps. The steps highlight the progression from initial situation understanding, to identifying and evaluating relevant reasoning threads, to applying synthesizing reasoning to the reasoning threads and resolving one's understanding, and responding/acting based on the resolved understanding. These steps may occur cyclically over time when a situation is being continuously monitored. This entire process is included in the scope of what is judged to be normative reasoning and behavior for the situation. For the entire process, mathematical and non-mathematical reasoning, and the use of knowledge of domain and situation/application detail apply. There are actions within the steps to seek out additional information to clarify or confirm one's understanding of the situation, including, for mathematical reasoning, to confirm that required assumptions hold for the situation.

Normative reasoning and behavior is presented in two forms within the framework. First, normative reasoning and behavior for a type of situation is agreed upon by expert consensus within a domain of application, and is accepted by the community. Proficient practitioners provide examples of this normative reasoning and behavior. Second, a personal norm for reasoning and behavior in the situation applies when an individual is not proficient in performing normative reasoning as settled by the expert consensus. For the person, there is a real risk of misapplying knowledge and reaping negative consequences, and so it is normative for the person to consider that risk in reasoning and behavior for the situation.

An example

Let us illustrate features of the framework, and of normative reasoning and behavior under uncertainty, with an example based on a real world situation. Consider the case of a hiring manager reviewing candidates for a job opening, with the purpose to determine which of the candidates to hire, or to hire none and keep looking. This situation involves uncertainty since the manager does not have complete knowledge of the candidates and how they may perform in the job. The situation involves risk, due not only to that uncertainty, but also because there are negative consequences to making a poor decision (such as hiring a candidate who ends up not performing well in the job, or who leaves the job after a short time, after company resources have been spent training him), and positive consequences to making a good decision (such as having a person exceptionally well suited to the job stay in the job, grow in the job, and serve the company well for years).

A first question is whether there is a normative standard for reasoning and behavior in this situation. Even though the situation of evaluating candidates for a job and making hiring decisions, is a common and important situation, there is no worldwide consensus on best practices for hiring people for jobs. However, there may be a consensus on such practices for a smaller community.

Let's say that Company A has developed a consensus on their hiring practices. For each candidate, they conduct a structured interview with the candidate and rate the interview; and they administer a test of job skills that are required for the job, and generate a skill performance profile for the person based on the test. To be hired, a candidate's job interview rating must fall within a specified acceptable range, and his skill profile must rate in a specified acceptable range for all required skills. If more than one candidate satisfies these conditions, the hiring manager judges, based on the interview and skill ratings, and his own judgment, the best candidate for the job. As defined, the hiring process involves heuristic reasoning. This is the hiring practice that has been in place for the past 5 years, and Company A may revise its practices in the future if improvements are identified and agreed upon. The company gets many of its job candidates from three educational/training institutions X, Y, and Z, and maintains historical records on the number and percentage of its hires that come from each of the three institutions. Over the past 5 years, new hires have been 25% from X, 50% from Y, and 20% from Z.

So, to follow the progression of the normative reasoning and behavior defined by Company A, using the steps highlighted in the cognitive framework, first, the hiring manager gets an initial situation understanding: he understands his typical goal to make a hiring decision, who the available candidates are this time, and the results of the structured interview and test of job skills for each of the candidates. Then the manager identifies and evaluates the relevant information to making the hiring decision (reasoning threads) for each candidate, considering each piece of information relevant to the decision. At this stage, the evaluation may lead to a candidate being eliminated from consideration because he does not meet the minimum requirements. Let's say that four candidates have acceptable ratings for the interview and the test of job skills. Now, synthesizing reasoning needs to be applied to consider all four remaining candidates and which is the best bet to hire. For this step, the company's normative process does not tightly constrain the decision, only saying that the manager is to consider the interview and skill ratings and also apply his own judgment. Consistent with these guidelines, the manager reviews in detail and compares the information for the four candidates. He eliminates two candidates who had mostly minimally acceptable ratings. Reviewing the full information on the two remaining candidates P and Q, the manager sees that one candidate Q has very high ratings for several skills, including a skill for which his other team members are not strong. He sees it as positive to be able to add someone strong in that skill to his team. In contrast, candidate P's ratings are all in the midrange. The manager decides to hire candidate Q.

A colleague, who knew that there were the two candidates P and Q in the final running for the job, suggested to the hiring manager that he hire candidate P, because P was from institution Y, which historically supplied 50% of new hires to the company, more than the 25% supplied by institution X, from where candidate Q had come. The hiring manager replied that there was no agreement in the company that that factor must be applied; and that he believed that one reason that institution Y provided more new hires was that more of their students were Spanish-speaking, which was a required skill for their jobs; and that candidate Q had that skill. The hiring manager also pointed out that, by basing hiring decisions on individuals' performance, and not on the institution where they were trained, the process was fair, not giving preference to candidates based on where they came from. The colleague listened with interest, and departed with a greater self-awareness of gaps in his own knowledge related to making a good choice for new hire.

The above example illustrates normative reasoning and behavior established by expert consensus, the use of heuristic reasoning and domain knowledge, and the richness of situation/application detail relevant to reasoning. In the example, there is mathematical information available on the base rates of new hires by educational/training institution, but that information is not required to be used in the process, and, in the example, reasons are provided for not using that information in the decision to hire an individual.

This hiring decision example structurally parallels examples that Tversky and Kahneman (1974) used to support their conclusion that people's use of heuristics, such as the representativeness heuristic, in reasoning under uncertainty, "lead[s] to systematic and predictable errors" (p.1131), such as ignoring base rates or prior probabilities of outcomes. In one of Tversky and Kahneman's examples, subjects were asked to judge the probability that a person has the occupation of engineer or lawyer, given a personality sketch as well as base rate information on the engineer vs. lawyer mix of the pool from which the sketch allegedly had been randomly selected. They found that subjects tended to use the base rate information if no personality sketch was provided; but tended to use only the personality sketch and ignore the base rate information when the personality sketch was provided to evaluate. The hiring decision example shows a situation where it is normative to ignore available base rate information. In a parallel fashion, subjects in Tversky and Kahneman's study ignore base rates in favor of focusing on the information provided about the particular individual being judged. Regarding synthesizing reasoning, Tversky and Kahneman (1974) presuppose that people should follow Bayes' rule, a straightforward number crunching, in making a judgment based on the combined evidence from the personality sketch and the base rate information. In contrast, in the hiring decision example, the synthesizing reasoning occurs in the domain of the real world, richly using situation detail, and is not just a numerical calculation in the realm of mathematics.

Implications for Educational Practice

The mathematics of probability and statistics provides a powerful tool for use in reasoning under uncertainty, in situations when it is applicable. There is also risk of misapplying the mathematics. A view of what is normative or appropriate reasoning under uncertainty, including normative application of probability and statistics, provides valuable guidance for the effective and adaptive reasoning and behavior in real world situations; as well as valuable guidance for educational practice regarding what should and what should not be taught. Such a framework for normative reasoning and behavior under uncertainty has been presented in this paper. Based on the framework, recommendations for educational practice in probability and statistics follow:

1. *Demonstrate the power of probability and statistics when used normatively/ appropriately, in different domains.* Provide real world examples that illustrate established normative standards for the application of probability and statistics in different domains, to show the power of probability and statistics when used appropriately. Use the established methods within the domain, which may integrate mathematical and non-mathematical reasoning, and provide the richness of detail involved in applying the methods.

2. *Use the domain of games of chance to teach probability and statistics.* Games of chance provided the original problems that mathematicians used in the first development of the math of probability and statistics (David, 1962; Gigerenzer et. al., 1989), and provide a well-established domain for the normative application of probability and statistics. Moreover, experience with the concrete instantiation of randomness in the common random phenomena associated with games of chance (e.g., the rolling of dice, and the blind drawing of balls from an urn), helps to build the mature understanding of random phenomena that is a foundation for understanding probability and statistics (Kuzmak, 2014).

3. *Don't teach students to ignore relevant domain knowledge and application/situation detail in situations involving uncertainty.* Appropriate use of probability and statistics is not just choosing math formulae that seem to fit a situation, and plugging in numbers, and calculating, regardless of application detail and whether required assumptions are satisfied. If one is teaching in a way that models that kind of behavior, then one is teaching students to reason inappropriately, specifically, to ignore domain knowledge and application detail that is relevant to appropriate reasoning. Statistics problem solving should regularly include analysis of the domain of application adequate to justify the applicability of the math to the situation, and to not merely assume its applicability.

4. *Explicitly teach the risks of misapplying probability and statistics.* To prevent students from misapplying probability and statistics, and from using information based on misapplication of others, explicitly teach common ways that probability and statistics can be misapplied, that they should avoid. Examples of such misapplication were provided in the above section on that topic, e.g., not ensuring a sample is representative of the population about which one wishes to make inferences, and not taking into account required assumptions when applying statistical techniques. Give examples of the negative consequences of misapplying probability and statistics, to emphasize the importance of avoiding such misapplications.

5. *Explicitly teach that the use of the math of probability and statistics is not always the best or normative approach to reasoning and behavior under uncertainty.* Give illustrations of normative reasoning under uncertainty in different domains, that use heuristic reasoning, and a combination of heuristic and mathematical reasoning. This provides further emphasis that the math of probability and statistics is a tool to be used when appropriate, which is not always. Give examples illustrating that there is not always an established normative approach to reasoning and behavior for a situation, and that there may be controversies over valid approaches, including controversies over how probability and statistics may apply.

Conclusion

The focus of this paper has been on providing a cognitive framework for viewing normative reasoning and behavior under uncertainty in real world situations, including situations involving risk. Such a framework provides a valuable reference for guiding what should be taught, and, in contrast, what should not be. Recommendations for educational practice in probability and statistics based on aspects of the normative framework have been provided in the section above. The framework for normative reasoning and behavior also provides a reference for evaluating whether particular examples of observed reasoning and behavior, identified as "errors" in the research literature, are rightly called so. Within the framework, a consensus of experts within the domain of application is the determinant of what is normative. The errancy of assuming that applying probability and statistics to a situation is normative, or assuming that every situation has an established normative response, has been noted.

The framework provides a fresh perspective on factors that are fundamental to normative reasoning and behavior under uncertainty, emphasizing the roles of mathematical and non-mathematical reasoning, and real world domain knowledge and situation/ application detail, and illustrating the roles of these factors with several examples rich in relevant detail. Within the framework, mathematics is viewed as a tool that may be applied appropriately or inappropriately, not a universal solution.

The risk of misapplying probability and statistics is identified as a significant risk at play in situations involving reasoning and behavior under uncertainty, and is illustrated through examples, discussion, and research findings. One faces this risk when one is not proficient at the normative reasoning and behavior agreed upon by domain experts. The framework recognizes that, in this case, it is a personal norm to consider the risk in one's reasoning and behavior in the situation.

References

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences, 4*, 317-370.

Chrissis, M., Konrad, M., & Shrum, S. (2011). *CMMI for development: Guidelines for process integration and product improvement* (3rd ed.) (pp.481-495). Boston: Pearson Education, Inc. http://cmmiinstitute.com/resources/cmmi-development-version-13

David, F. (1962). *Games, gods and gambling: The origins and history of probability and statistical ideas from the earliest times to the Newtonian era.* London, UK: Charles Griffin & Co. Ltd.

Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*, 44-63.

Garfield, J. & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*, 592-596.

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of chance*. Cambridge, UK: Cambridge University Press.

Gigerenzer, G. & Todd, P. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp.3-34). New York: Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, U.K.: Cambridge University Press.

International Standards Organization (ISO) (2009). *ISO 31000: 2009, Risk management – Principles and guidelines.* Geneva, Switzerland: International Organization for Standardization. http://www.iso.org/iso/home/standards/iso31000.htm

International Electrotechnical Commission (IEC) (2009). *IEC 31010: 2009, Risk management – Risk assessment techniques.* Geneva, Switzerland: IEC. http://www.iso.org/iso/catalogue_detail?csnumber=51073

Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine, 2*, 696-701.

John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524-532.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus, and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, U.K.: Cambridge University Press.

Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582-591.

Kuzmak, S. (2014). What's missing in teaching probability and statistics: Building cognitive schema for understanding random phenomena. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. http://icots.info/9/proceedings/

Lopes, L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 626-636.

Lopes, L. & Oden, G. (1991). The rationality of intelligence. In E. Ellis & T. Maruszewski (Eds.), *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science* (pp. 199-223). Amsterdam: Rodopi.

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359-1366.

Tishkovskaya, S. & Lancaster, G. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education, 20*(2). Retrieved from www.amstat.org/publications/jse/

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Good Models and Good Representations are a Support for Learners' Risk Assessment

Laura Martignon
Ludwigsburg University of Education, Germany
Sebastian Kuntze
Ludwigsburg University of Education, Germany

Abstract: When learners have to make sense of risky situations, they can use mathematical models and representations which facilitate successful risk assessment. Based on theoretical considerations on the benefits of specific models and specific representations in such contexts, we present empirical findings of a study which examined whether students use such models and representations in their risk assessment. We conclude that the availability of adequate models to learners may help them gain transparency when facing risks and thus foster their decision-making.

Keywords: risk, representations, models, assessment, decision-making.

Introduction

If learners are unable to model risk, dealing with risk is hindered – learners may view the loss of resources as wholly unpredictable. They will hardly have any criteria for assessing expectations and may develop excessive fear becoming risk averse – or – conversely – become risk prone and engage carelessly in high-risk endeavors.

Human life is characterized by risk: most decisions and actions have a risky side to them. The difference between the ways we deal with risks today and the way our ancestors dealt with them is that we now have theoretical support to analyze risks. Simply defining risk as the possibility of losing a (valuable) resource with a certain probability provides a framework that facilitates assessing risk. A typical example is a gamble: someone offers us to throw a coin and if heads turns up we gain, say 10 dollars. Yet if tails turns up we have to pay 5 dollars. What is the risk here? Should we accept the gamble? Observe that we lose 5 dollars with a probability of 0.5 (our risk) and gain 10 dollars also with a probability of 0.5 (our benefit). Thus it should be considered worthwhile to accept the gamble. We say that a situation (an experiment) is risky when at least one of the events with strictly positive probability produces a loss of resources (e.g. time, health, energy, money). Most decisions we make daily have risky sides, which we should be able to evaluate (Gigerenzer, 2014; Kurz-Milcke, Gigerenzer & Martignon, 2008; Martignon & Krauss, 2009). Risk analysis is today a highly developed sub-discipline of decision analysis and requires sophisticated mathematical treatment. Yet the basic elements for getting along in everyday life are simple and can be grasped even by young children.

Representations that foster risk assessment

In this section we illustrate two basic tools for risk assessment: distinguishing between absolute and relative risk and understanding conditional probabilities. We often hear that the use of a medicine may reduce the risk of a certain disease by, say, 50%. How should we interpret this information? Actually, there is something missing in this piece of information. In fact it communicates a relative risk without specifying the absolute risk.

Consider the following "true story": In the mid-1990s, the British press reported the results of a study that women who took this contraceptive pill increased their risk of thromboembolism by

100%. *Thromboembolism* means blockage of a blood vessel by a clot and can lead to fatal strokes. Hearing the bad news, thousands of British women panicked and stopped taking the pill, which led to a wave of unwanted pregnancies. But what did the study in fact show? Out of every 7.000 women who did not take the pill, one had thromboembolism, and out of every 7.000 who took it, this number increased from one to two. An absolute risk is communicated with mention of the reference grade, whereas a relative risk only considers the gain or loss with respect to a specific element of this grade. Understanding the difference between absolute and relative risks should be part of the school curriculum.

Good representations for an elementary introduction of relative and absolute risk reduction in Primary School

The dynamical webpage www.eeps.com/riskicon presents convenient representation tools for instructing school students on the difference between absolute and relative risk. Figure 1 is extracted from that webpage. The simulation lets you explore risk reduction. A number of people have had bike accidents. Half of them were wearing helmets. The relative risk reduction when wearing a helmet is of 50%. Yet the absolute risk reduction is of 2 out of 10 to 1 out of 10. The sliders at the right hand side of the iconic illustration can be used to vary the size of the sample, the probability of injury when riding without a helmet and the risk reduction provided by the use of a helmet.

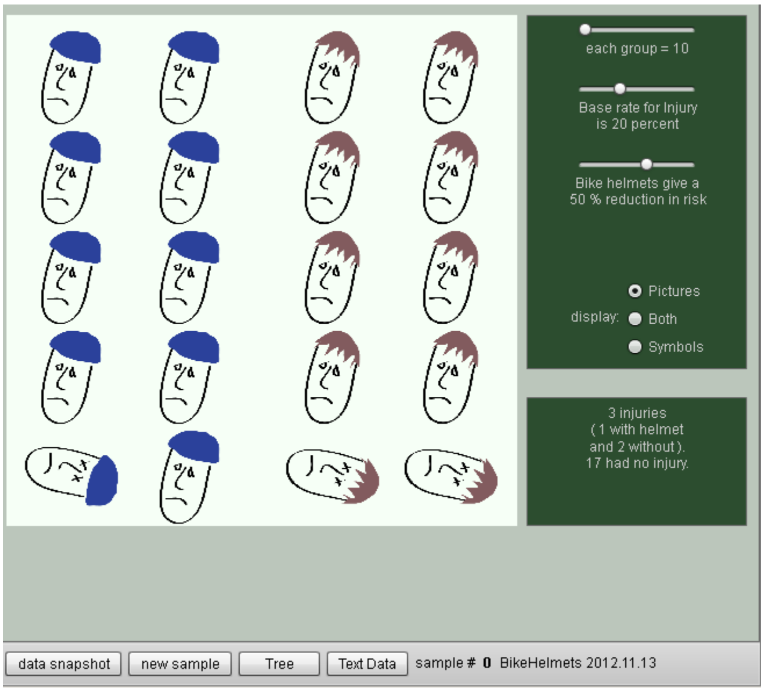


Figure 1. Absolute/relative risks with and without helmets

Another fundamental tool for reckoning with risks is conditional probability. Understanding conditional probabilities is essential when we are faced with decisions on our health: symptoms or tests tell us, with a certain probability, whether we suffer from a certain disease or not. How do we assess the probability of having a disease given that we have the corresponding symptom? This is a conditional probability. It is usually simpler to assess the probability that the symptom is present given that we have the disease (causal direction) than the other way round, namely to assess the probability of our having the disease when the symptom is present (diagnostic direction).

In order to compute, say, the validity of a test or symptom for diagnosing a disease we have to reason the Bayesian way. This means, that given the base rate or probability of a disease, its sensitivity (probability that a patient tests positive when she has the disease) and its specificity (probability that a patient tests negative given that she does not have the disease) we have to compute the probability of a patient having the disease given that she tests positive. Severe mistakes in this type of reasoning are made not just by lay people but also by experts which, according to the research in the Heuristics and Biases program of Kahneman and Tversky (1982), is mostly due to a base-rate neglect. Here the ABC Group led by Gerd Gigerenzer has proven how useful it is to resort to so called natural frequencies and imagining a population of, say, 1000 people (Gigerenzer & Hoffrage, 1995): In this imagined the number of persons out of these 1000 who have the disease corresponds exactly to the base rate, while the proportion of true positives out of the grade of people with the disease corresponds exactly to the sensitivity and the proportion of true negatives out of the grade of people without the disease corresponds exactly to the sensitivity. Thus if the base rate or prevalence of the disease is 10% then exactly 100 of the 1000 people imagined have the disease. If the sensitivity of the test is 8% then exactly 80 of the 100 people with the disease test positive, and if the specificity is 90% then exactly 891 out of 990 without the disease will test negative. In a strict mathematical sense the so-called natural frequencies correspond to expected frequencies: one expects that 100 out of the 1000 imagined people have – on average – the disease. Eliminating the hurdles of probabilities, like randomness and variability, expected frequencies transport the framework of the Bayesian task back to the simple arithmetic of proportions. Such a setting facilitates reasoning immensely (Gigerenzer & Hoffrage, 1995).

Good representations for an elementary introduction to Bayesian Reasoning in Primary School

Traditional developmental theory, as initiated by Piaget and Inhelder in the second half of the last century, suggests that children do not become proficient at making probabilistic inferences until age 7 (Piaget & Inhelder, 1975). Nevertheless, more recent research has shown that very young children are capable of performing simple probability calculations when task demands are reduced (Zhu & Gigerenzer, 2006). In general, many of Piaget's postulates on children's logical and probabilistic intuitions have had to be revisited and modified. For instance in their book "The early Growth of Logic in the Child" Inhelder and Piaget (1964 [1959]: 101) reported an experiment in which they showed 5- to 10-year-old children pictures, of which 16 were flowers and 8 of these 16 flowers were primroses. The children were asked several questions. One of them was "Are there more flowers or more primroses?" Only 47% of 5- to 7- years old children gave answers in accord with grade inclusion − that is that reflected un understanding that the flowers were more numerous because they included the primroses as a subset. Among 8-years olds, however, a majority (82 percent) gave answers consistent with grade inclusion. This kind of experiment, according to Inhelder and Piaget, indicated that children acquire logical intuitions even without instruction. Of course, this experiment has probabilistic connotations as well: flowers are more probable than primroses. It is important to note, in fact, that the step towards a proportional statement like "8 out of 16 flowers are primroses" is quite natural, once children learn to understand and formulate proportions. Here again, Piaget's pioneering results have been revised by recent experiments on children's early estimates of proportions for simple categorization.

The important discovery in this context has been (Martignon & Krauss, 2007, Multmeier, 2012) that the representation of information plays a fundamental role in children's and adults' understanding of proportions and probabilities. Multmeier investigated children in second and in fourth grade and divided his samples out of second and fourth classes in two groups. Both groups had to solve six tasks like the following:

> *In a small and faraway fairyland village there are 10 inhabitants: 2 princesses and 8 mermaids. Of the 2 princesses, 1 wears a crown. Two of the 8 mermaids also wear a crown. If I tell you I met an inhabitant of the fairyland village and mention this inhabitant was wearing a crown, would you bet it was a princess?*

In one group of children, second graders (7 years) and fourth graders (9 years), had to solve the task based on the texts of the tasks only. In another group, children, second graders and fourth graders, had to solve the tasks based on texts and iconic visualizations, like the following, which corresponds to the task above, on princesses and mermaids:

Figure 2. Iconic representation of a fairyland village

The following graph exhibits the results of this experiment:

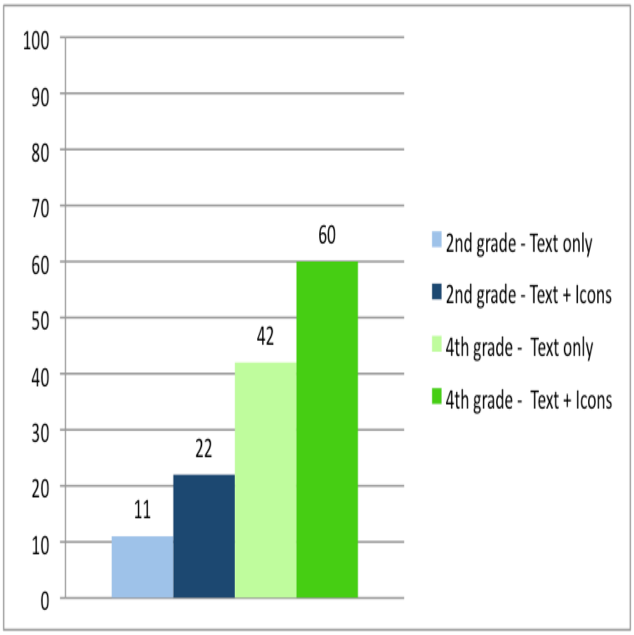


Figure 3. % correct of second graders (n = 91, 45 text only) and fourth graders (n = 85, 42 text only)

The results show that children in fourth grade have much stronger intuitions for proto-Bayesian thinking (i.e., reasoning on conditional frequencies and their inversion) than children in second grade yet they also show that iconic representations foster proto-probabilistic intuitions in both classes. They encourage instruction in fourth grade on proto-Bayesian reasoning by means of the systematic use of appealing iconic representations.

A dynamic webpage for instructing children in primary and secondary school

Selecting the village of mermaids and princesses as an inviting environment for teaching first steps of conditional probability and Bayesian Reasoning the dynamic webpage www.eeps.com/riskicon introduced above, also presents illustrations like the following:

Figure 4. Icon Array to represent a population of mermaids and princesses

In Figure 4 ten stylized inhabitants of a fairyland village are represented: 2 princesses and 8 mermaids. The user has a menu of possibilities: by using sliders he can vary the base rate of princesses among the fairy folk. She can also vary the "sensitivity" of "crown", i.e., the probability, that a princess wears a crown, and also the specificity of a crown, namely the probability that a mermaid does not wear a crown. She can click on the "tree" button to obtain the "double tree" describing this situation, say, of 20 fairy folk, 6 princesses, 4 with a crown and 14 mermaids, of whom 6 wear a crown:

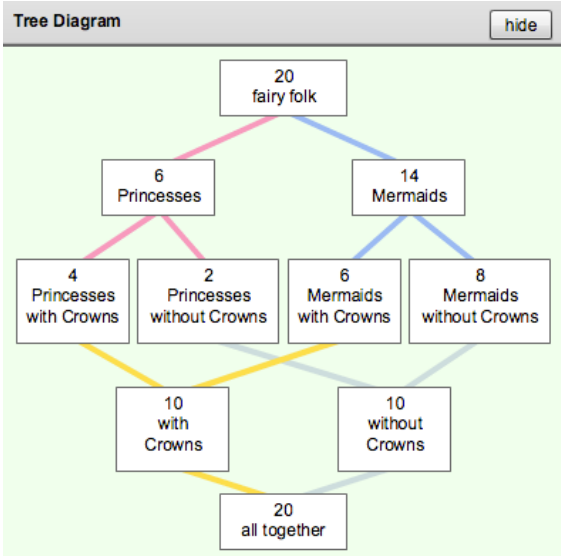


Figure 5. The double tree describing natural frequencies of a fairy folk environment

The double tree represents both directions, the direction from "princesses" to "crowns", which can be called the causal direction ("being a princess may cause wearing a crown"), and the diagnostic direction, from "crowns" to "princesses" ("wearing a crown can be a sign that one is a princess"). The double tree facilitates reasoning not just from cause to effect, but also in the inverse direction, from effect to cause, i.e., Bayesian reasoning with natural frequencies. The dynamic webpage also presents a HIV environment, designed for school students in ninth or tenth grade, where again the user can vary many parameters and visualize the corresponding double trees.

Reasoning with natural frequencies and double trees can be seen as a form of proto-Bayesian reasoning, which constitutes a good heuristic for Bayesian Reasoning with probabilities. First elements of proto-Bayesian reasoning can be conveyed to children as early as in fourth grade, not just with good iconic representations (mermaids and princesses, as in Figure 1) but also with concrete hands-on materials (Till, 2014).

In conclusion, we can think of many situations in which iconic representations foster the understanding and the assessment of risk not just of adults but of school students, young and old. We have also stressed the usefulness of dynamic representations with appealing materials.

Modelling and Comparing Risks

Modelling risks in several steps, as we shall see, makes these risks understandable. In analogy with the modelling cycle by Blum & Leiss (2005), the situational context of a risky situation first needs to be structured and/or simplified, in order to identify relevant characteristics of the situation, in order to determine the resource in danger and in order to establish criteria for describing and evaluating the potential loss of such a resource. This step does not yet require the use of mathematical models or quantification. Translating the situational model into a mathematical model is the subsequent step, which does require mathematical concepts, such as probabilities or expectations. Using a mathematical model for handling the assessment of a risk may lead to a mathematical result, which has to be interpreted against the background of the situation. The result interpreted in the context of the model then has to be validated. At this stage, additional criteria may come into play which could lead to the need of an improved or more refined mathematical model.

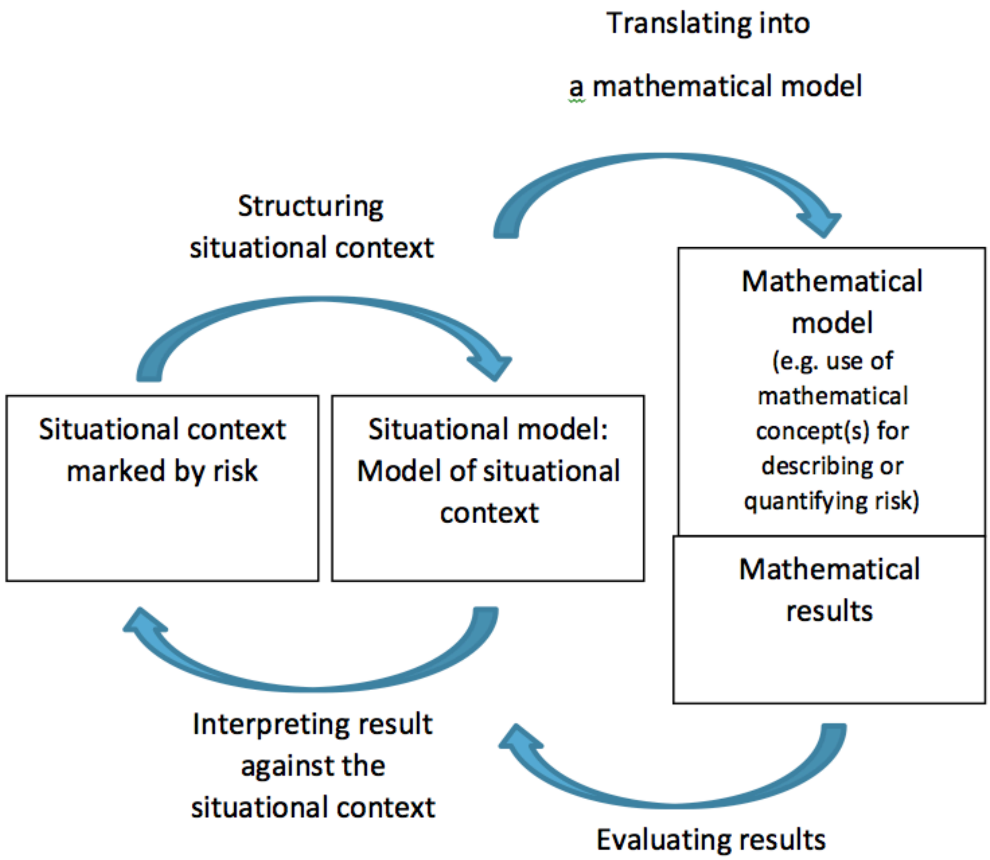


Figure 6. Modelling risky situations (see process models in Blum & Leiss, 2005)

Comparing Risks Playing Ludo

We will present here results of a study we carried through in Baden-Württemberg on the statistical perception and understanding of risk in school: we tested students of fourth grade in primary schools and ninth grade in "Real Schulen" (technical-track secondary schools), as well as undergraduates. We describe only one item of our test devoted to the Ludo game. This task allows us to perform an analysis of school students' use of representations and models for assessing risks.

We begin by introducing the Ludo Game:

Ludo: Players take it in turn to roll a single die. A player must first roll a six to be able to move a piece from the starting area onto the starting square. In each subsequent turn the player moves a piece forward 1 to 6 squares as indicated by the die. When a player rolls a 6 the player may bring a new piece onto the starting square, or may choose to move a piece already in play. Any throw of a six results in another turn. If a player cannot make a valid move she or he must pass the die to the next player. If a player's piece lands on a square containing an opponent's piece, the opponent's piece is captured and returns to the starting area. A piece may not land on square that already contain a piece of the same color. Once a piece has completed a circuit of the board it moves up the home column of its own color. The *player must throw the exact number to advance to the home square. The winner is the first player to get all four of their pieces onto the home square.*

We proposed the following task in fourth-classes and in ninth-classes in and near Stuttgart:

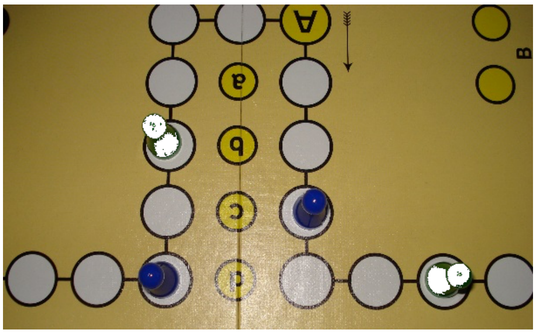


Figure 6. A risky situation for “White” in the Ludo game

The students had the instruction: *Consider two players Black and White. In one position, Black is two fields behind White, and at another position Black is three squares behind White. Assume it is White’s turn. He rolls a “one”. Which piece should White move? Which move is riskier?*

For an adequate modelling in this example, learners may argue that the probability of rolling a “three”, which is risky for the case that the piece on the left is moved, is smaller than the probability of rolling “two or four”, which is risky for the case when the piece on the right is moved. In an alternative adequate model, learners could consider the number of equally probable possibilities. It is hence interesting, whether learners of different age succeed in modelling the risks in the Ludo situation adequately. However, an additional perspective connected with the issue of modelling is of interest, namely how learners represent risks verbally, that is, how they communicate about risks. Even if it has to be taken into account that basic abilities in mathematics and statistics are often necessary for elaborated risk assessment, communicating about risks requires above all dealing with these abilities in a context-related way. In the realm of risk communication, two quality levels of students’ reasoning have been described (Engel, Kuntze, Martignon & Gundlach, 2010):

•Level 1: identifying risks verbally, describing the possibility of loss of a resource

•Level 2: Weighing/comparing risks, identifying decision alternatives and comparing them against the background of risk

Higher levels could also make sense, for instance the use of relative versus absolute frequencies when communicating about risks, the ability to translate from one to the other of these types of frequencies, the ability to establish the completeness of information provided by texts which inform about risks. We will use the example in Figure 6 (“Ludo task”) in order to illustrate the two Levels mentioned above. The resource at risk is, in the case of White, the pieces on the board, which should not be captured by the other player. Identifying this resource and describing the danger of losing it corresponds to Level 1 of risk communication. In case the risks associated with different decision alternatives of White are effectively compared, Level 2 is reached. These levels are treated independently from the decision made in the end. The quality of risk communication is the issue here.

Research Interest

Against the background outlined in the previous section, our preliminary study focused on the following research questions: *How do students assess risks in the context of the Ludo task? Do they succeed in modelling the risks in this situation and how do they communicate about these risks?*

Design of the Study

In order to investigate these research issues, the Ludo task was presented to primary students, secondary students and undergraduate students. This study was carried out in the framework of the project “RIKO-STAT”. We report about results from 385 fourth-graders from 21 classes at primary schools) and from 549 ninth-graders from 21 classes at technical-track secondary schools.

The choice of the Ludo task is due to its connection to risk and risk reduction. Observe that the risk of White losing a piece is reduced by ½ if the piece at the left hand side advances one field. Here ½ is the relative risk reduction, while the absolute risk reduction is of 1 piece. Besides the Ludo game belongs to the typical table games played by young students in Germany.

Results

In the following, we will first present some answers from students of different ages. This will illustrate not just our coding of the students’ answers but also provide some qualitative insight into their reasoning. We will then turn to overview results on code frequencies in these subsamples. We begin by presenting Noemi’s answer. She is a primary student (4th grade) who gives only the short answer displayed in Figure 7.

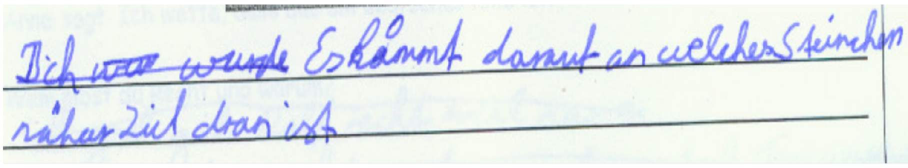

Ich ~~war~~ ~~wurde~~ Es kommt darauf an welches Steinchen
näher Ziel dran ist

Figure 7. Noemi’s answer (4th grade).

Noemi writes: “It depends upon which piece is closer to its final goal”. In her answer, Noemi does not describe the risk, instead she mentions the criterion of being closer to the goal as the only basis for making the decision. This is obviously a generally meaningful aspect of the Ludo situation. However, in this case, this criterion is not sufficient for guaranteeing a good model of the risks at stake here. For this reason, this answer was coded as non-adequate modelling. Figure 8 now shows the answer given by Lara (9th grade).

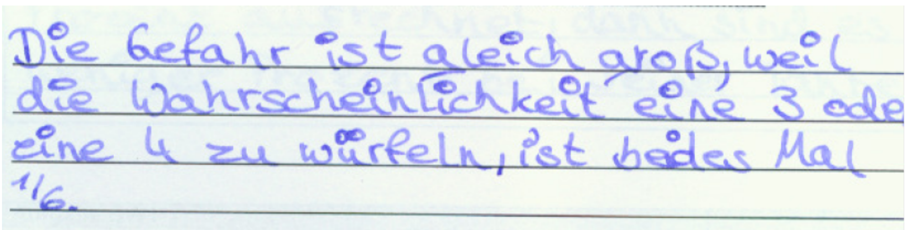

Die Gefahr ist gleich groß, weil
die Wahrscheinlichkeit eine 3 ode
eine 4 zu würfeln, ist beides Mal
1/6.

Figure 8. Answer by Lara (9th grade)

Lara says that the danger is the same, because the probability of rolling a 3 or rolling a 4 is 1/6 in each case. From Lara’s answer, we can see that she knows that the probability to roll a 3 or a 4 (she meant a 2 or a 4) is each time 1/6. However, she appears not to have remarked that in this case, the probability of rolling a 2 or a 4 has to be compared with the probability of throwing a 3 in order to compare the risks in the probability model she has chosen. Lara mentions a “danger”, but she does not explain in detail what the danger consists of, with respect of the possible decisions. Describing the risks explicitly might have helped her to see and distinguish the possibilities of losing one piece with “2 or “4”. This could have supported a modelling adequate to the problem situation. Figure 9 now exhibits an answer, in which the risk is verbalized rather indirectly.

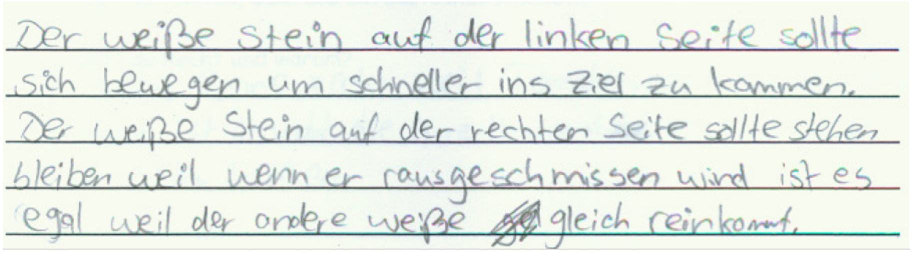

Der weiße Stein auf der linken Seite sollte sich bewegen um schneller ins Ziel zu kommen. Der weiße Stein auf der rechten Seite sollte stehen bleiben weil wenn er rausgeschmissen wird ist es egal weil der andere weiße gleich reinkommt.

Figure 9. Answer by Dennis (9th grade)

Dennis' answer is: "The white piece on the left should move in order to reach the goal sooner. The white piece on the right should stay because if it gets caught this plays no role as the other white piece will reach the goal immediately." Dennis thus describes risks of pieces being caught only for the alternative that the piece at the right hand side is moved. We may, yet, assume that he is generally aware of this type of risk. Moreover, Dennis compares the risks associated with two decision alternatives for White. It is however probable that Dennis interprets the grey fields a, b, c, and d in the middle as the goal of the white pieces. This interpretation is unfortunately wrong. In fact, the task description specifies that the pieces shown are the last ones in the game and so two other pieces must already have reached the goal. Moreover, merely arguing in terms of the distance to the goal is in general not sufficient, because the probabilities of loss of resources have to be properly considered. The answer was hence coded as not exhibiting adequate modelling. The following answer by Katja (9th grade), illustrated in in Figure10, uses relationships between frequencies of possible outcomes when rolling the die.

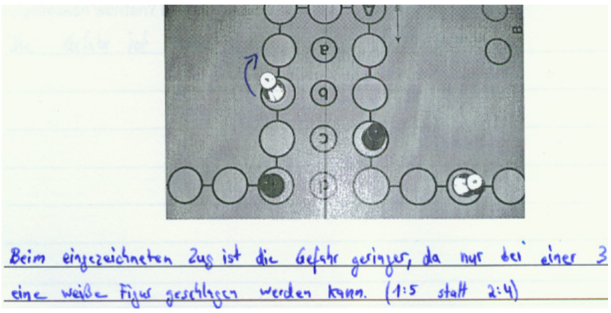


Figure 10. Answer by Katja (9th grade)

Katja's answer was: "The white piece on the left should move in order to reach the goal sooner. The white piece on the right should stay because if it gets caught this plays no role as the other white piece will reach the goal immediately." Katja's answer shows a very short verbal identification of the risk, including comparisons between decision alternatives. Moreover, she uses an adequate modelling of the situation, by counting positive and negative possibilities and comparing their proportions ("1:5 instead of 2:4") In this case, the use of the model appears to be closely connected with reasoning on the comparison of the decision alternatives. As Figure 11 now shows, even primary students can be able to model risks adequately and to describe risks in order to compare them reasonably.

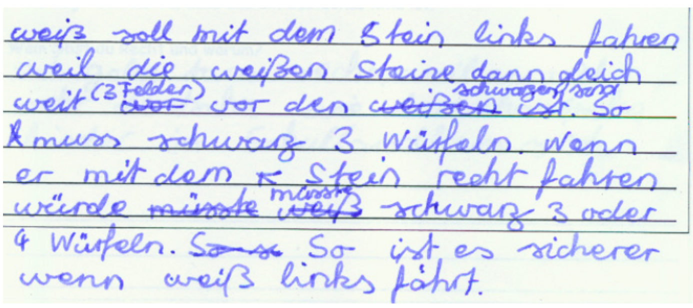

weiß soll mit dem Stein links fahren weil die weißen Steine dann gleich weit (3 Felder) ~~weg~~ vor den ~~weißen~~ schwarzen sind. So muss schwarz 3 würfeln. Wenn er mit dem Stein recht fahren würde ~~müsste~~ müsste ~~weiß~~ schwarz 3 oder 4 Würfeln. ~~So~~ So ist es sicherer wenn weiß links fährt.

Figure 11: Answer by Lasse (4th grade)

Lasse answers: "White should move the piece on the left, as the white pieces are then equally far (3 fields) from both black pieces. Thus Black must roll a 3. If he moved the piece on the right hand side, Black would have had to roll a 3 or a 4. Thus it is safer for White to move the piece on the left hand side."

Lasse (4th grade) models mainly by considering all possible cases in a structured way, by asserting that they are basically equally probable. The small mistake (“3 or 4” instead of “2 or 4”) obviously does not imply that Lasse did not model the problem adequately. Looking at the way he communicates about risks, we have here a case in which even if risks are neither identified verbally nor described in detail, it is obvious that Lasse compares different decision alternatives against the background of risk considerations. Lasse’s answer was hence coded as having reached Level 2 of risk-related communication.

The overview results shown in Figure 12 suggest that it was not easy for the students of both age groups to model risk adequately.

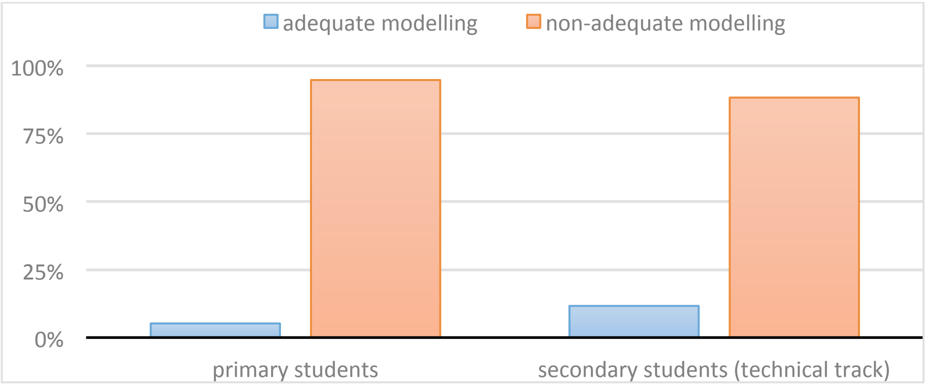


Figure 12: Relative frequencies of codes related to modelling

The relative frequencies show that even 9th-graders (technical track secondary schools) did not exhibit relevant conceptual knowledge to the extent that would have enabled them to model risks successfully.

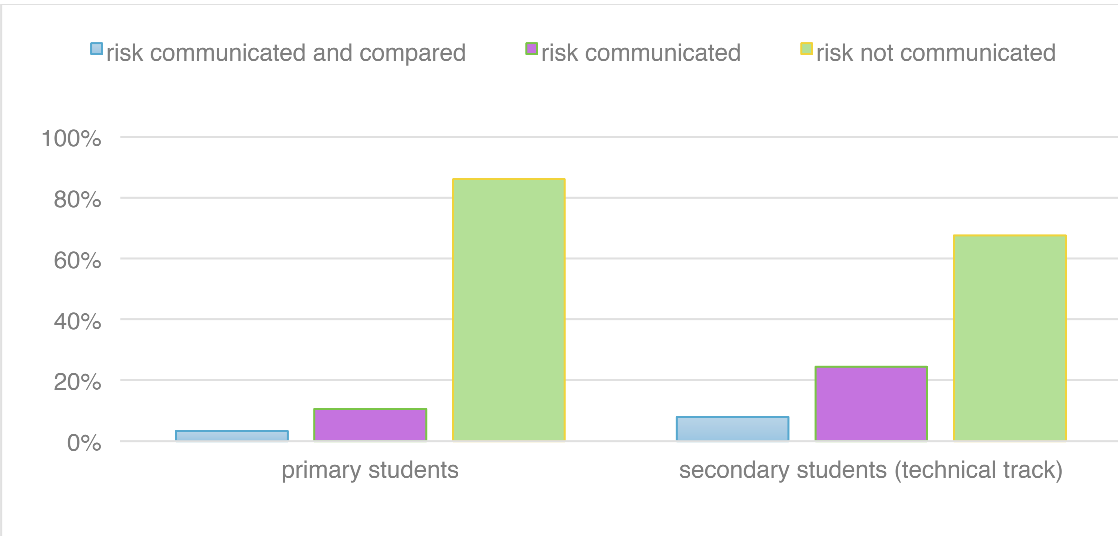


Figure 13. Relative frequencies of codes related to communicating risks

The descriptive results of the code frequencies in Figure13 indicate that a majority of students from both groups did not give verbal descriptions of the risks. Only a minority of the students gave comparisons of risks and different decision alternatives.

Discussion and Conclusions

The results of this study suggest that school students are in need of elementary strategies for modelling risky situations as well as they need strategies for verbalizing and communicating about risks. To express by means of language what the danger of losing a valuable resource amounts to, and in which cases such a loss of a resource may happen is an extremely useful strategy. It does not only support an understanding of a problem situation, but it can contribute substantially to an orientation and to a structured consideration of the risky situation. Mathematical models may subsequently emerge from these strategies for communication about risk. Our proposal here is to prepare school students in the understanding of risks and on the communication of risks by means of the systematic use of transparent, iconic representations.

References

Blum, W. & Leiss, D. (2005): How do students and teachers deal with mathematical modelling problems? The example “Sugarloaf”. In ICTMA 12 Proceedings (pp. 222-231). Chichester: Horwood.

Gigerenzer, G. (2002). Reckoning with risk: Learning to live with uncertainty. London: Penguin Books.

Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: frequency formats. Psychological Review, Vol 106 (2), 425-430

Kahneman, D., Tversky, A. & Slovic, P. (1982) Judgement under Uncertainty: Heuristics and Biases. Cambridge University Press

Kuntze, S., Engel, J., Martignon, L. & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge – Empirical research in the project RIKO-STAT. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS).

Kuntze, S., Lindmeier, A., & Reiss, K. (2008). „Using models and representations in statistical contexts” as a sub-competency of statistical literacy – Results from three empirical studies. *11th International. Congress on Mathematics Education (ICME 11). http://tsg.icme11.org/document/get/474*.

Kurz-Milcke, E., Gigerenzer, G. & Martignon, L. (2008). Transparency in Risk Communication. Graphical and Analog Tools. Ann. N.Y. Acad. Sci., 1128, 18-28.

Martignon, L., & Krauss, S. (2009). Hands on activities with fourth-graders: a tool box of heuristics for decision making and reckoning with risk. International Electronic Journal for Mathematics Education, 4(3), 117–148.

Martignon, L. & Kurz-Milcke, E. (2006). Educating children in stochastic modelling: Games with stochastic urns and coloured tinker cubes. In A. Rossmann & B. Chance (Eds.), Working cooperatively in statistics education. ICOTS 7 – Salvador de Bahia. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/17/C443.pdf, June 22, 2009

Multmeier, J. (2012) Representations facilitate Bayesia reasoning: Computational facilitation and ecological design revisited. Berlin: Freie Universität Berlin, 2012 (Dissertation)

Piaget, J. & Inhelder, B. (1975). The Origin of the Idea of Chance in Children. Routledge and Kegan Paul. Translation of original work (1951).

Till, Ch. (2014) Fostering Risk Literacy in Elementary School. Mathematics Education, 9(2), 83-96

Zhu, I. & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. Cognition, 98, 287-308.

Teaching Risk in School

Andreas Eichler

University Kassel, Germany

Markus Vogel

University of Education Heidelberg, Germany

Abstract: Although risk is an important topic for society it is seldom addressed when teaching statistics and probability. In this paper we refer to this discrepancy identifying three obstacles for teaching risk in school regarding the mathematical and the situational aspect of risk. Based on two educational constructs, i.e. probability literacy and modelling, we discuss existing approaches for teaching risk in school and propose two strategies for promoting risk as a valuable issue for students based again on the distinction of the mathematical and situational aspect of risk.

Keywords: risk, risk literacy, probability literacy, modeling.

Obstacles for Teaching Risk in Schools

Although risk is understood as a "part of our everyday lives" (Erickson, 2010) and risk is "a hot topic" (Spiegelhalter, 2014) that should be taught in school (cf. Gigerenzer, 2013), this hot topic seems to play at most a minor role in school, particularly if mathematics teaching is regarded. We approach this only seemingly paradox with the following situation (Latten, Martignon, Monti, & Multmeier, 2011, p. 21) that was developed for mathematics teaching:

> Is it more risky to possess an Alsatian dog (German shepherd) than a Chihuahua? It is known that four out of nine Chihuahuas and three out of ten Alsatian dogs bite their owner at most once.

The specific task was to compare the "risk" of possessing an Alsatian dog or a Chihuahua. The intended answer was that it is more risky to possess an Alsatian dog, because the result of a bite from this dog breed is considerably worse than a bite from a Chihuahua, although the probability of a Chihuahua's bite is bigger than that for a bite from an Alsatian dog. This task includes a first obstacle for teaching risk in school. The obstacle is not the dog-task itself, but the amount of different answers. Following different authors who work on risk issues, we found mainly three possible different answers:

1. As mentioned above, the answer of the task developers is that the bite of the Alsatian dog represents the bigger risk. This answer is based on the definition of risk as the arithmetical product of a measure of the dis-utility represented by a random event and the probability p of this event, i.e. *risk* = p x *measure of the dis-utility of an event* (Latten et al., 2011). A disadvantage of this definition is the problem of measuring a dis-utility (c.f. Kent, Pratt, Levinson, Yogui, & Kapadia, 2010). Of course a bite from a dog represents a dis-utility. However it is not clear how to measure the loss of healthiness based on a dog's bite. Further, the mentioned arithmetical product does not exist in certain cases. For example a bite from the white killer shark (in Steven Spielberg's film Jaws) results in a person's death. Thus, if we define the dis-utility of a person's death as infinite the risk of a bite from the killer shark is infinity. However, the definition of risk

discussed above is often used in the literature and is close to the theory of the Subjective Expected Utility, a theory of human decisions (Savage, 1954).

2. In contrast, Gigerenzer (2013, pp. 39-40) defines risk as a probability of an event that is measurable. The event could represent both good luck and bad luck or rather “nice or nasty” outcomes (Spiegelhalter, 2011). A disadvantage of this definition is the unclear distinction between risk and probability. Although Gigerenzer (2013) discusses a lot of examples that intuitively could be assigned to the construct of risk (e.g. financial risks, diseases, games of chance), nearly every probability of a random event could be understood as a risk since nearly every random event represents in a certain sense a positive or a negative outcome. For example, both the probability of a bite from an Alsatian dog and the probability of a 6 when throwing a die represent a risk. The definition of Gigerenzer (2013) is close to the (economical) risk definition of Keynes (1921), who defines the Laplacean probability and the frequentistic probability as risk, but a subjective probability as (a person’s) uncertainty.

3. On a first view, the third “definition” includes the demand not to define the term risk: “Risk is a strange concept. Different disciplines have tried to define it precisely, but perhaps it is better to be informal” (Spiegelhalter, 2011, p. 17). On a second view, risk is implicitly defined as a situation’s characteristic consisting of a pair of a probability and the related random event that mostly could be understood as a bad event. Also this definition has the difficulty that there is no clear distinction between a situation with risk and a situation of uncertainty (without risk). For example, a pair of a dog’s bite and the related probability could be a characteristic of a risky situation as well as the pair of a 6 when throwing a die and the probability of getting a 6. However, we follow in this paper this third definition that is as a specific situation’s characteristic also used by the International Risk Governance Council (IRGC, 2005; Boeckelmann & Mildner, 2011).

Besides these three definitions, there are further aspects of the construct of risk that we mostly neglect in this paper. For example, many researchers refer to the human influence on risk (Slovic, 2010) or rather the human inability to cope with risks (e.g. Kahneman, 2011). Further, another aspect of risk is the distinction between calculable situations and not calculable situations. The latter form of situations is discussed as unknown unknowns (e.g. Boeckelmann & Mildner, 2011; Gigerenzer, 2013) or black swans (Taleb, 2007).

The plurality of definitions of the concept of risk and the fact that not all definitions are appropriate to confine the concept of risk to already existing concepts, e.g. probability, is a first obstacle for teaching risk in school, particularly, for teaching risk as a part of mathematics. Thus, the plurality of definitions and the absence of clear differentiation from other existing and well defined concepts seem to impede a clear conception of a risk curriculum. E.g. Kent et al. (2010) state that “risk is a difficult topic for teachers and the detailed questions about what and how to teach remain (for mathematics especially) largely unanswered”.

A further obstacle for teaching risk in school could be caused by the situations that are connected with risk in the related literature and, particularly, in daily media. Thus, risky situations are often nasty, complex and refer to data for which the process of data generation is hard to understand. For example, a lot of risky situations include nasty outcomes, i.e. include a heavy loss of healthiness (diseases, accidents) or a heavy loss of financial resources (economics, games of chance). However, it should be scrutinized whether situations referring to cancer or death by accident are suitable for teaching mathematics in school. Further, risky situations that are adequate for teaching mathematics in school must be graspable for students in a mathematical sense. Finally, the demand of a data driven curriculum (Moore, 1997) includes that data and data generation are crucial issues for statistics and probability teaching. By contrast, for example, it seems to be a big challenge to unfold the generation of data that are necessary to judge the risk of using different forms of transportation (walking, cycling, motorbike, car; Spiegelhalter, 2011).

Although there are possible obstacles for teaching risk in schools regarding the *mathematical aspect* of risk in terms of an appropriate definition as well as referring to the *situational aspect* of risk, the demand to improve students' risk literacy as a crucial skill of "informed decision making" (Till, 2014) in situations of uncertainty is, nevertheless, important for teaching mathematics in school. For this reason, we integrate firstly the mathematical aspect and the situational aspect of risk into overarching models of teaching mathematics or teaching statistics and probability. Afterwards we allocate existing approaches for teaching risk or risk literacy to these educational models. Finally, we propose different examples aiming to reduce the obstacles of teaching risk in school.

Probability Literacy and Risk Literacy

Gal and Garfield (1997, p. 3) give a meaty fundament to reflect what is important to be reached as an overarching goal at the end of the curriculum: students should become "[…] informed citizens who are able to […] comprehend and deal with uncertainty, variability, and statistical information in the world around them, and participate effectively in an information-laden society". Referring to this fundament, the issue of decision-making under uncertainty is close to the aim of risk literacy as mentioned above. As with this brief summary of the overarching goals, other well-known conceptions for overarching goals of teaching statistics and probability seem to integrate risk – partly without mentioning the term risk, or risk literacy. For example, Franklin et al. (2005, p. 1) emphasize the ability to deal with "decisions in our personal lives". Of these conceptions we particularly focus on Gal's definition of statistical literacy (Gal, 2002) and probability literacy (Gal, 2012) since both conceptions include explicitly both the mathematical aspect of dealing with decisions of uncertainty and the situational aspect. For example, Gal (2002) refers to a statistical and mathematical knowledge base (mathematical aspect) and to world knowledge and critical skills (situational aspect) as main aspects of statistical literacy. Similarly, Gal (2012) describes "probability literacy" as a part of the broader construct of statistical literacy. He defines probability literacy as follows: "The ability to access, use, interpret, and communicate probability-related information and ideas, in order to engage and effectively manage the demands of real-world roles and tasks involving uncertainty and risk." (Gal, 2012, p. 4). For the sake of emphasizing the risk aspect Till (2014) speaks of "risk literacy" when he refers explicitly to this definition of Gal (2012). For probability literacy Gal (2012) defines two points of view as being constitutive:

- The internal view addresses the students' learning within the pure mathematical world by understanding and applying logic, abstract concepts, proofs, formulas, graphs, etc. In general, it is about dealing with mathematical objects and mathematical representations or computing probabilities in specific probability models. The internal view could represent the mathematical aspect of risk.
- The external view means basically that teaching probability should (beyond the internal view) comprise the role of probability-related ideas when for example talking with students about sampling or testing hypotheses. At its broadest level the role of probability-related ideas should

be seen within their potential to contribute to the educational goal that the students can become informed citizens. In this broad sense, the external view represents the situational aspect of risk.

Both views are intertwined. Thus, when focusing on teaching risk within this perspective of Gal's definition of probability literacy those problem situations, which are feasible for the students with regard to both the situational context and the situation's underlying mathematical structure, become interesting.

Probability literacy as well as risk literacy fit also the main aspects of modelling as an overarching conception of mathematics education (Blum, Galbraith, Henn, & Niss, 2007). Thus, risk in its mutual relationship (cf. Eichler & Vogel, 2012) between real-world affairs comprising real situations and empirical data, and mathematical-world affairs comprising probabilities could be seen as a specific form of modelling (fig. 1) referring to the following four modelling phases:

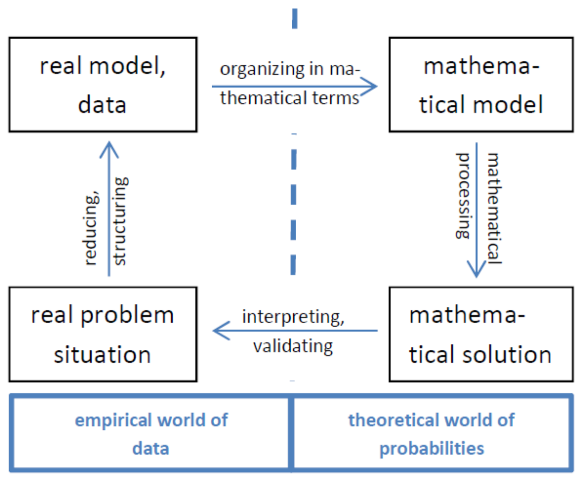


Figure 1. A model of modelling

• A real problem situation concerning risk has firstly to be structured and reduced to the relevant information.

• The so-called "real model" (Blum et al., 2007) is the basis for applying mathematics originating from the theoretical world of probabilities, for example relating two random events and the concept of probability to each other by asking for a certain conditional probability within a risky situation.

• Based on these mathematical models some mathematical results can be derived like, for example, fixing a concrete probability as a risk.

• These mathematical results have to be translated in the real-world, e.g. for a decision in a risky situation.

The underlying educational idea of modelling also refers to teaching risk based on probabilistic concepts and methods. Firstly, it is about students' ability to access real-world problems with their available mathematical concepts and methods: "Our mathematical concepts, structures, ideas have been invented as tools to organize the phenomena of the physical, social and mental world." (Freudenthal, 1983, ix). Secondly, it is about students' ability to deepen their understanding of mathematics and to develop new mathematic concepts and methods (c. f. Klieme, Neubrand, & Lüdtke, 2001, 143).

Both ideas can be interpreted in the light of Gal's (2012) points of views mentioned above: The internal view of educational goals of probability education can be seen in relation to the second idea, in concrete, the deepening of understanding of probabilistic concepts and methods; the external view can be seen in relation to the first idea, in concrete, enhancing intellectual accessibility to risky problem situations or situations comprising uncertainty applying available knowledge about probabilistic concepts and methods.

Existing Conceptions of Teaching Risk

Independently from the obstacles that we identified above, there exist several educational approaches for teaching risk in schools. We report some of these ideas referring to our theoretical framework including

- the differentiation of teaching risk and teaching statistics (or probability), and
- ideas for dealing with the internal view (mathematical aspect) and with the external view (situational aspect) of risk.

Not surprisingly, researchers that address the question of risk literacy, predominately rank teaching risk among teaching statistics and probability (e.g. Latten et al., 2011; Martignon & Monti, 2010). In addition, Kent et al. (2010) explicitly mention risk also as a part of science teaching in UK.

Regarding the internal view, an explicit risk curriculum is missing. However, Till (2014) following Martignon and Krauss (2009) mention proportions, conditional probabilities and expected values as a basis for risk literacy. Implicitly, Gigerenzer and his team at the Harding Center for risk literacy (https://www.harding-center.mpg.de/en/) give notes for possible subjects of a risk curriculum when listing technical terms referring to the rubric of health information that are a basis for risk literacy. These technical terms encompass concepts that are common for a probability curriculum except of the concept of distribution, e.g. probabilities (and frequencies), conditional probabilities, Bayes' formula, independence. A substantial extension of the probability curriculum is represented by the crucial distinction of absolute risk and relative risk (e.g. Gigerenzer, 2013).

A main focus of the approaches to teaching risk literacy, which we have quoted in this paper, is on the interface between the internal view (mathematical aspect) and the external view (situational aspect) of risk in terms of fostering understanding and interpreting risk situations by using specific representations of the situations. The tree diagram with natural frequencies is a main part of these representations (e.g. Gigerenzer, 2013) since research yielded its potential for solving tasks in risky situations (Sedlmeier & Gigerenzer, 2001). Other representations that are often discussed are for example fast and frugal trees (e.g. Laskey & Martignon, 2014) or colored cubes in a hands-on activity (e.g. Kurz-Milcke & Martignon, 2006). Following Spiegelhalter and Gage (2014), who on their part took the concept from Gage (2012) the use of physical models (the colored cubes) and multiple representations is a core for teaching probability (and risk).

Although there are situations of risk with nasty outcomes in the literature (e.g. Kent et al., 2010, referring to a "Painometer" or Eichler & Vogel, 2014 referring to HIV), there are a lot of attempts to transfer main elements of risk literacy to situations in which the risky event is far from being a threat for students. This is especially the case if risk literacy is addressed in primary schools. For example, Till (2014) uses the situation of snails biting plants in a garden (task referring to proportion) or focuses on the situation that a person has to judge the gender of a person when knowing a specific characteristic of the person like the hair length (Bayesian formula). Another example for secondary schools is given by Erickson (2010), who addresses, in a playful way, insurance problems. Although we did not find a clearly given rationale to use situations to promote risk literacy that are not nasty it seems to be an undeclared demand to mostly avoid these nasty situations for teaching risk in schools.

The mentioned conceptions for teaching risk in school stem from educational literature. However, from a teacher's perspective, the first and foremost question in preparing a course concerning risk is the question of available material in text books. Kapadia (2008) states with regard to an analysis of mathematical text books used in England: "It is certainly true that the majority of teachers make links to real-life situations, indeed they often occur in text-books. But the lack of reference to risk is very important in the context of the current world." A corresponding screening of a small sample (n=13) of well-known mathematical text books for secondary school in Germany yields a similar result: 307 tasks were taken into account as touching different contextual themes in the field of making decisions under uncertainty and handling risk in a broad sense. For example these tasks include questions about tests concerning production quality and production errors, questions about elections and predicting probable outcomes of elections, questions about the principle of flights overbooking including calculations of possible losses, about elementary insurance problems, and about the diagnostic power of different health tests. These tasks address mathematical competencies going beyond those mathematical competencies like proportional reasoning, dealing with expected values and understanding conditional probabilities which Martignon and Krauss (2009) call "tool box for decision-making and reckoning with risk". The main mathematical topics of the investigated upper-secondary tasks were: conditional probabilities (including Bayes formula), distributions (binomial and normal), testing hypotheses and confidence intervals. With regard to the tasks' contextual background it was interesting to see that in less than 5% of the regarded 307 tasks the word "risk" was used explicitly. Although we do not raise the claim of representativeness our finding meets the statement of Kapadia (2008) quoted above.

A Proposal for Dealing with Obstacles for Teaching Risk in Schools

We identified three obstacles for teaching risk in schools. The first obstacle refer to the mathematical part of risk and, more precisely, is based on the absence of a mathematical definition of the construct of risk and on a difficult delimitation of risk to existing concepts of probability. A first proposal for promoting risk as a part of mathematics teaching includes aspects

- to understand risk as a pair including a mathematical aspect and a situational aspect,
- to identify the difference between risk and an uncertain situation that we do not understand as risk, as only being related to situational aspects but not to the mathematical aspects. Thus, risk is inextricably intertwined with interpreting the real world, but does not exist solely in the mathematical world;
- to acknowledge that referring to situational aspects or rather modelling is one of more goals of mathematics teaching. Thus, both the internal view and the external view in their broadest sense are important issues for improving probability literacy.

Secondly, another obstacle refers to emphasizing risky situations including nasty outcomes as a part of mathematics teaching. Our proposal of promoting the teaching of risk in school is to consider more explicitly the situational aspect of risk and to vary more explicitly risky situations with nasty outcomes and situations that are structurally equivalent but have no nasty outcomes.

Finally, the third obstacle, i.e. the nearly complete absence of the construct of risk in text books might be a consequence of the two obstacles summarized above. For this reason, we neglect the third obstacle in the following, but discuss an example that illustrates our proposal for promoting teaching risk in school.

The Contextual Side: Varying Situations

We differentiate three kinds of problem situations (Eichler & Vogel, 2009; Eichler & Vogel, 2014):

• Problem situations which contain all necessary information and represent a concept of probability (or statistics) like a conditional probability in a strongly reduced context or virtual context we call *virtual problem situations.*

• Problem situations which demand analyzing a situation's context that is more "authentic" and provide a "narrative anchor" (Cognition and Technology Group at Vanderbilt, 1990), but still provide demands that do not aim to reconstruct real world problems we call *virtual real world problem situations.*

• Problem situations that include the aim to reproduce real problems of a society we call *real world problem situations.*

These problem situations differ with regard to their level of the situation's reality. Varying available material along these different levels of reality is based on the idea of reflecting on a certain probabilistic concept and/or method in a continuum between the empirical world of society, nature and techniques which comprises the whole complexity of reality including risks and the theoretical world of pure mathematics, in specifically the world of probabilistic concepts and methods. We give an example of the three worlds referring to the same mathematical structure.

A *real-world problem situation* that is both strongly related to the term of risk (e.g. Gigerenzer, 2013) and connected to a very nasty outcome is an HIV-test. The interpretation of the result of an HIV-test is further doubtless a highly relevant topic of public health debates all over the world. The core problem can be captured by tasks like the following which we found in different variations in our screening of available materials in German text books (cf. as well Eichler & Vogel, 2010):

> How reliable is an HIV-test?
>
> In Germany the prevalence of HIV infection is estimated to be 0.1%, according to official sources. An HIV test which is usually conducted as a multi-stage process can provide a high level of certainty regarding the presence of the infection.
>
> At the first stage such an HIV test has the following characteristics: sensitivity and specifity. The sensitivity of an HIV test gives a correct indication of infection in 99.9% of all cases tested. The specificity of the test means that 0.5% of cases will be wrongly identified as having HIV, although infection is not present.
>
> If a person's first test yields a positive result, how certain can this person be the he or she is actually infected?

Research results of the working group of Gigerenzer (2004) yielded that even for professional physicians it is not easy to grasp such conditional probability problems concerning apprppriate interpretation of medical test results. In terms of the modelling cycle presented above, a possible modelling structure could be as follows:

• The real problem situation here is the question off the reliability of an HIV test given the situation as described in the task.

• Building the so-called real model would mean extracting the relevant information given in the text (i.e. base-rate; sensitivity and specificity of the test), relating them to each other and setting them in relation to what is asked for in the task, a measurement for the certainty of really being infected given a positive test result.

• Transforming the information of the real-world into the mathematical world can be described as three-stepped: Firstly, the relevant information is to be captured within terms of (conditional) probabilities, here it would be (with T+ standing for a positive test result and HIV- standing for not being infected): base-rate P(HIV) = 0.001 (which implies P(HIV-) = 0.999), it is given here proactively only because of its need in the Bayes formula), test sensitivity P(T+|HIV) = 0.999 and test specificity P(T+| HIV-)=0.005. In a second step the requested information, i.e. the probability of being really infected has also to be expressed mathematically by using conditional probabilities: P(HIV|T+). Third, these probabilities have to be applied to the Bayes-formula:

$$P(HIV|T+) \quad = \frac{(P(T+|HIV)\cdot P(HIV)}{(P(T+|HIV)\cdot P(HIV)+(P(T+|HIV-)\cdot P(HIV-)}$$

$$= \frac{0.999\cdot 0.001}{0.999\cdot 0.001+0.005\cdot 0.999} = 0.17$$

• This mathematical result, i.e. 17%, has to be transformed back into the real-world and interpreted in the light of the given information: Following the findings of Gigerenzer's research most of the people having heard from a positive HIV-test result would probably not think that the risk to be really infected is that low.

Of course, maybe there are students who know this topic very well because of being touched with the underlying problem by knowing people being HIV-infected within their family or their circle of friends. In this case, it is questionable from a pedagogical point of view whether one should tackle such an example in this class if the teacher decides to tackle it he or she has to be very careful. However, it is possible to integrate the same mathematical structure in an (uncritical) *virtual real world problem* that is shown below (c.f. Eichler & Vogel, 2014).

Students are playing a game along the following rules (c.f. Riemer, 1991): There are small groups of students each with one referee. The referee has two dice, an ordinary unbiased die representing an equal distribution of probability and a cuboid shaped die with the distribution: 1–0.05; 2–0.10; 3–0.35; 4–0.35; 5–0.10; 6–0.05. The dice are covered for the other members of the referee's group. The referee has covertly chosen one of the dice and throws it repeatedly. The other students try to find out which die the referee has chosen by processing information about the outcomes of each throw and by estimating step-by-step the probabilities for the die being covertly thrown.

From a mathematical perspective, the die task is structurally equivalent to the HIV test (fig. 2), particularly, if one result of throwing the die is given.

HIV-task	Die-task
available information prevalence: P(HIV+) = 0.001 thus, P(HIV-) = 0.999 sensitivity of the test: P(T+|HIV+) = 0.999 specifity of the test: P(T+|HIV-) = 0.005	available information prevalence: P(N) = 1/2 (normal die) thus, P(C) = 1/2 (cuboid die) sensitivity of the test: P(6|N) = 1/6 specifity of the test: P(6|C) = 0.05
calculation: $P(HIV+\|T+) = \frac{P(T+\|HIV+)\cdot P(HIV+)}{P(T+\|HIV+)\cdot P(HIV+)+P(T+\|HIV)\cdot P(HIV-)} = \frac{0.999\cdot 0.001}{0.999\cdot 0.001+0.001\cdot 0.005} = 0.17$	calculation: $P(N\|6) = \frac{P(6\|N)\cdot P(N)}{P(6\|N)\cdot P(N)+P(6\|C)\cdot P(C)} = \frac{1/6\cdot 1/2}{1/6\cdot 1/2+1/20\cdot 1/2} = 0.77$

Figure 2. Structural equivalence between the HIV-task and the die-task

The risk regarding this virtual reality could be defined. Consider for example the situation where the correct conclusion regarding which die was thrown when showing a certain number e.g. 6 yields a win of 5 Euros, but the false assumption costs one billion Euros. This would be a very risky situation compared to the situation where a correct or incorrect answer would have no consequences outside the game.

The same mathematical structure could also be integrated in *virtual world problem situation.* As mentioned before, we suggest reducing the complexity of the situation by keeping the mathematical structure. Here in the context it could be a short comic-story like the following one:

For many years Kurt has had a rainmaker friend who is able to make it rain by singing his special rain song. Because it is very exhausting the rainmaker only sings on 10% of all days. Gradually year by year, the rainmaker loses his magic power increasingly and he now succeeds only 80% of the time in making it rain when he sings his special rain song.

One day in May, Kurt had his first rendezvous with Karla, the beauty queen of his class. Kurt chose May for his date because he knows that in May it rains only 10% of the days (without accounting for the rainmakers influence) and thus, the probability of having a bad hair day seems to be low. All went perfectly, but just right at that moment when he wanted to kiss her it started to rain, and she leaves not being kissed because of having her hairstyle totally destroyed.

Kurt is frustrated and he suspects his rainmaker friend. On what probability for being guilty is Kurt justified in suspecting the rainmaker?

With regard to the purpose of concentrating closer on the mathematical structure a virtual problem situation has the advantage of being freely constructible for the mathematical narrator. Thus, exaggerations and simplifications are possible as well as variations of story influencing parameters which can be used to emphasize structural aspects of the underlying mathematical model like the Bayesian formula. Regarding risk, a potential nasty outcome in this situation is always a virtual nasty outcome. A further advantage of virtual world problem situations is that students accepting such a simplified virtual world situation could refer more directly to the situation's underlying mathematical structure, because understanding and processing the contextual information is not as hard. The modelling cycle turns out to be useful for describing the solutions process's structure of a virtual problem situation of decision–making under uncertainty by transferring its terms and considering them in parallel (fig. 3).

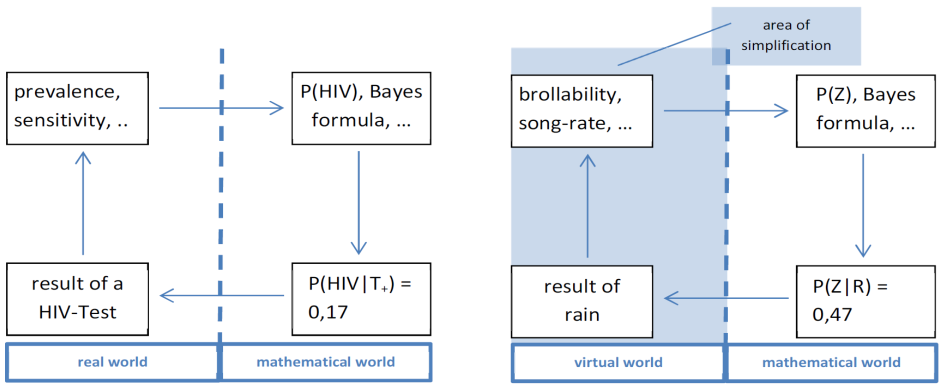


Figure 3. Analogies between solving the HIV-problem and the rainmaker-problem

In this parallel point of view, it turns out that the problem solving structure of a virtual problem situation of uncertainty like the rainmaker-problem can be seen as being similar to the modelling structure of the HIV example discussed above. A decisive difference is that a serious real-world problem situation is exchanged for a simplified virtual real-world problem situation, thus, the students have to what can be called "re-model" the virtual situation prepared by the teacher.

Again, from a mathematical perspective, the rainmaker task is structurally equivalent to the HIV test (Figure 4).

HIV-task	Rainmaker-task
available information prevalence: P(HIV+) = 0.001 thus, P(HIV-) = 0.999 sensitivity of the test: P(T+|HIV+) = 0.999 specifity of the test: P(T+|HIV-) = 0.005	available information prevalence: P(Z) = 0.1 thus, P(Z) = 1 – 0.1 = 0.9 sensitivity of the test: P(R|Z) = 0.8 specifity of the test: P(R|Z) = 0.1
calculation: $P(HIV+\|T+) = \frac{P(T+\|HIV+)\cdot P(HIV+)}{P(T+\|HIV+)\cdot P(HIV+)+P(T+\|HIV)\cdot P(HIV-))}$ $= \frac{0.999\cdot 0.001}{0.999\cdot 0.001+0.001\cdot 0.005} = 0.17$	calculation: $P(Z\|R) = \frac{P(R\|Z)\cdot P(Z)}{P(R\|Z)\cdot P(Z)+P(R\|Z-)\cdot P(Z-)}$ $= \frac{0.8\cdot 0.1}{0.8\cdot 0.1+0.1\cdot 0.9} = 0.47$

Figure 4. Structural equivalence between the HIV-task and the rainmaker-task

At the end, we claim a psychological argument for considering the variation of situations including potentially a risky situation: Referring to situated learning theory Brown, Collins, & Duguid (1989) argue that in general, learning and cognition are fundamentally situated. Following this approach learning situations (HIV, dice, rainmaker) do not need necessarily to be real in the sense of what we call real-world problem situations (see above). It is not the level of reality but the level of the perceived authenticity and realism of a learning situation containing multiple possible perspectives on the problem (cf. Mandl, Gruber, & Renkl, 1997) which turns out decisive for coping with the situation. The internal logic structure of the situation must be consistent, traceable and comprehensible. By following the goal to prevent the so-called inert knowledge, the research group Cognition and Technology Group at Vanderbilt (1990) developed and applied in practical oriented research the theory of anchored instruction. The central characteristic of this approach is its narrative anchor by which the interest of the learners should be produced and the problem should be embedded in the context so that situated learning can take place. Thus, within a pedagogical interpretation of these theoretical approaches of situated learning and anchored instruction and a basic implication of cognitive load theory (Chandler & Sweller, 1991) the learning about structural mathematical concepts and methods like proportional reasoning or probabilistic argumentation is expected to be facilitated when we reduce the complexity of the problem situation with should students should cope. This includes virtual worlds like for example being given in tales, narratives or comic stories.

The mathematical side: Understanding structures

When concentrating on the mathematical aspects of teaching the Bayes formula the formula's structure comes into the foreground but not only the probabilities resulting from calculations with the formula. The student will not understand what the formula $\mathrm{P(A|B)} = \frac{\mathrm{P(A \cap B)}}{\mathrm{P(B)}} = \frac{\mathrm{(P(B|A) \cdot P(A)}}{\mathrm{(P(B|A) \cdot P(A) + (P(B|A-) \cdot P(A-)}}$ mathematically means by "only" calculating one example after another in classroom. For teaching purposes the mathematical relationship of the formula-combined probabilities becomes important with regard to static and dynamic aspects of variation. To reach this by teaching the fundamental idea is not to focus on the single representation of the Bayes formula but, by following Spiegelhalter and Gage (2014) to transform it into multiple representations which open up different perspectives on the underlying mathematical structure. It is necessary to transform the Bayes formula into different representations.

As already mentioned above, a very common and often used representation is the tree diagram, either with probabilities or by following Gigerenzer (2013) with natural frequencies. In case of our rainmaker problem this would appear as is shown in figure 5.

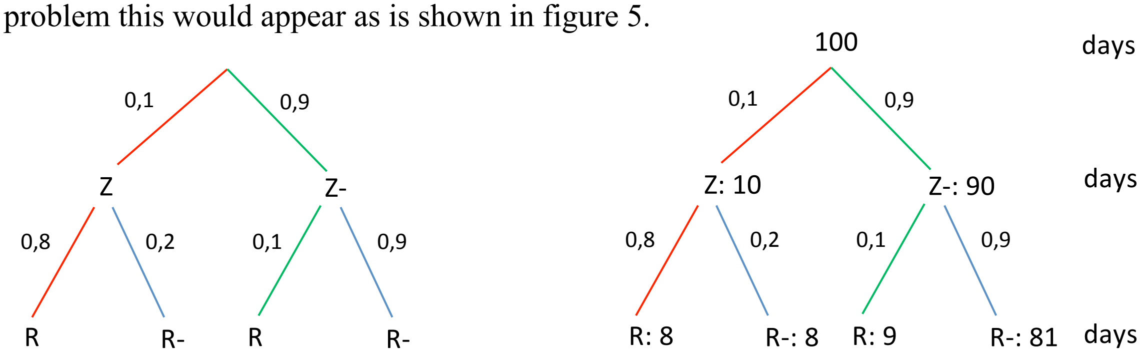


Figure 5. Tree diagrams with probabilities and natural frequencies concerning the rainmaker problem

The structural relationship to the Bayes formula could be represented by the following equation (fig. 6), the last "equation" visualizes those paths which are interesting for calculation by using colors.

$$P(Z|R) = \frac{P(Z \cap R)}{P(Z \cap R) + P(Z - \cap R)} = \frac{\text{probability for singing and rain}}{\text{probability for rain}} =$$

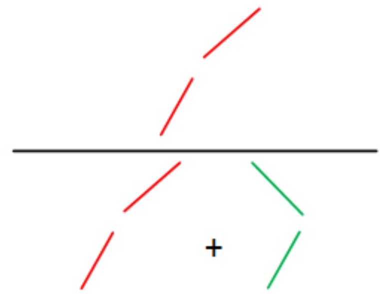


Figure 6. Relationship between Bayes formula and tree diagram

If there were natural frequencies used instead of probabilities the solution as well as the calculation of the problem becomes much easier for most of the people according to research findings of Sedlmeier & Gigerenzer (2001). However, the paths which have to be related to each other do not change, whether absolute frequencies were used instead of probabilities nor the values changed, for example for examining the changing proportions by varying the probability for singing with P(Z)=0,1; P(Z)=0,5; P(Z)=0,9. The tree visualization is static within this point of view (fig. 6).

For emphasizing the dynamic point of view we propose to use a unit square. A computer-based version of the unit square appears in outlines of the screenshot of figure 7: It is dynamically linked with the Bayes formula using the same colours as before for highlighting connections between corresponding elements.

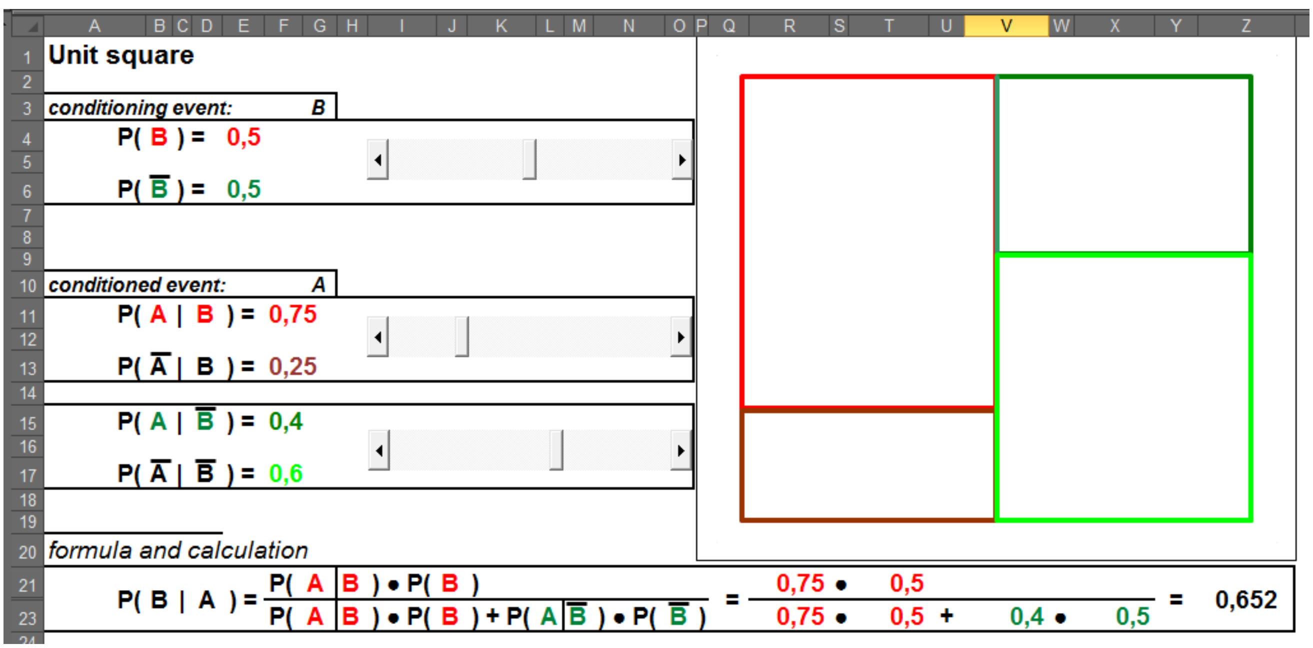


Figure 7. Computer-based learning environment concerning the Bayes formula

By using such a computer-based learning environment the students can explore the formula's structure, which is an important goal for mathematics teaching, particularly, when the mathematical aspect of probability literacy is regarded. In figure 8 the meaning of changing the base-rate is displayed within the visualisation of the unit square concerning the context of the rainmaker-problem.

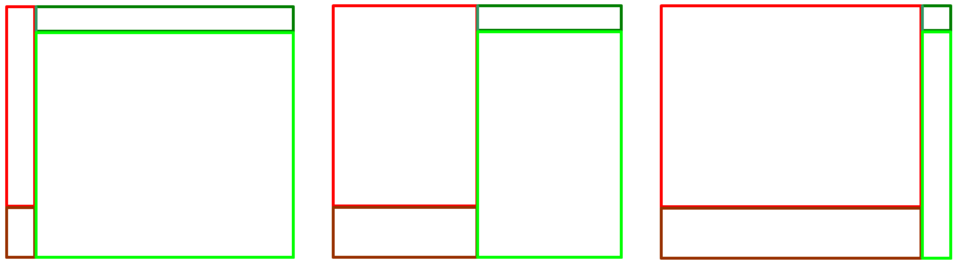


Figure 8. Influence of changing base-rate within the unit square (P(Z)=0,1; P(Z)=0,5; P(Z)=0,9)

The meaning of these changes with regard to a corresponding visualized version of the Bayes formula (cf. Eichler & Vogel, 2010) result in what can be seen in figure 9.

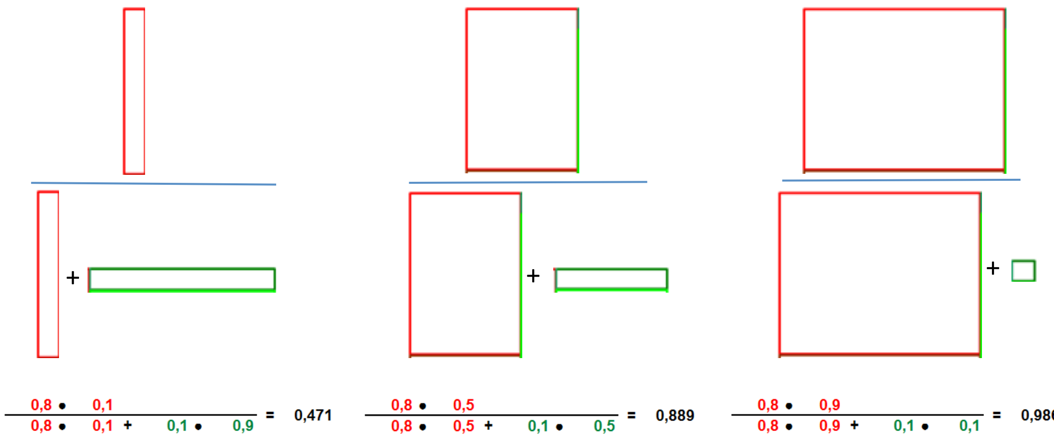


Figure 9. Bayes formula visualized by the unit square

In this point of view, the influence of varying the base-rate of rain-singing to the result could become more obvious by displaying the proportions of corresponding areas of the unit square. The unit square contains, at least partially, graphical elements of the mathematical structure of the Bayes formula but the tree diagram does not. For this reason, we prefer the unit square when focussing on teaching the mathematical understanding of the Bayes formula. For only finding a solution the tree diagram with the natural frequencies might be easier for most of the people outside the classroom.

Of course, this kind of visualization can also be applied to real-world problems like the HIV problem; here varying the base-rate would correspond to judging the test results in countries with different estimations of prevalence (cf. www.unaids.com) as well as to virtual real-world problems like the gambling situation mentioned above, in which varying the base-rate would mean changing the referee's preference for one of the two given dice, the normal one or the cuboid one. The unit square visualization focusses on mathematical structure but not the context.

Discussion and Conclusion

Why is risk a hot topic for society at large but plays a minor role in mathematics teaching in school? In this paper we claim that this fact could be provoked by obstacles for teaching risk in schools regarding the mathematical aspect and the situational aspect of the construct of risk. For this reason, it is a main issue of this paper to propose strategies to overcome the mentioned obstacles.

The first strategy concerns reducing contextual complexity by changing from the real world to the virtual real world (die task) or even to the virtual world (rainmaker). For example, this change means reducing the significance of the situation's risk because risk could matter in the world of games of chance, but does not matter personally within the virtual world. For example, the risk of Kurt is no risk for the reader of the comic. Serious risk is an issue of the real world but not the virtual world. This is at the core of the underlying pedagogical idea of changing from the real world to the virtual world: In the virtual world the students can think about the virtual context and the underlying mathematics without being seriously personally concerned. When they have learned about the structure and the contextual meaning in a virtual problem situation or in a game of chance they can transfer their knowledge to reality and thus, think about risk in its serious meaning concerning reality, like e.g. in the HIV problem.

This variation principle of tasks and problem situations can be applied to other mathematical concepts like confidence intervals, testing hypotheses or modeling with normal or binomial distributions. By changing different mathematical concepts as well as reality levels of problem situations, the experienced students can learn about the difficulties concerning different interpretations of what is meant if someone speaks of "risk" whenever they enter the real world.

Another main aim of this paper was to make a distinction between applying a heuristic to solve a real problem (the tree diagram with natural frequencies) and applying a heuristic to understand the structure of a mathematical problem (the unit square). Whereas the tree diagram with natural frequencies might be the optimal strategy for persons outside mathematics classes at school when faced with a problem like the HIV-test, the unit square seems to be more appropriate for yielding a mathematical and conceptual understanding of the Bayesian formula and the impact of the parameters to the arithmetical results (c.f. Eichler & Vogel, 2010; Böcherer-Lindner & Eichler, 2015).

In the end, we hold the view that acknowledging the perspective of teaching mathematics considering both the mathematical as well as the situational aspects of risk could be a sound strategy for making a hot topic for the society into a substantial topic within the teaching of statistics and probability.

References

Blum, W., Galbraith, P., Henn, H.-W., & Niss, M. (Eds.) (2007). *Modelling and applications in mathematics education. The 14th ICMI study.* New York: Springer.

Boeckelmann, L. & Mildner, S.-A. (2011). Unsicherheit, Ungewissheit, Risiko [Uncertainty, incertitude, risk]. *SWP-Zeitschriftenschau 2*, 1-8.

Böcherer-Lindner, K. & Eichler, A. (2015). Vergleich konkurrierender Visualisierungen zum Bereich der bedingten Wahrscheinlichkeiten [Comparison of different visualisations of conditional probabilities]. *Beiträge zum Mathematikunterricht*. Retrieved from http://www.mathematik.tu-dortmund.de/ieem/cms/de/home/bzmu_home.html.

Brown, J. S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Cognition and Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*(6), 2-10.

Eichler, A., & Vogel, M. (2009). *Leitidee Daten und Zufall [Key idea of data and chance]*. Wiesbaden: Vieweg+Teubner.

Eichler, A. & Vogel, M. (2010). Die (Bild-)Formel von Bayes [The (picture) formula of Bayes]. *PM - Praxis der Mathematik in der Schule, 52*(32), 25-30.

Eichler, A. & Vogel, M. (2012). Stochastik – fit für die Zukunft [Stochastics – fit for the future]. *PM - Praxis der Mathematik in der Schule, 54*(48), 2-9.

Eichler, A. & Vogel, M. (2014). Three Approaches for Modelling Situations with Randomness. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistics thinking* (pp. 75-100). Dordrecht: Springer.

Erickson, T. (2010). Exploring risk through simulation. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics.* Lublijana, Slovenia: International Statistical Institute and International Association for Statistical Education. Retrieved from http://iase-web.org/documents/papers/icots8/ICOTS8_8C3_ERICKSON.pdf.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer, R. (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. Alexandria, VA: American Statistical Association.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures.* Dordrecht: Reidel.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review, 70*(1), 1-51.

Gal, I. (2012). Developing Probability Literacy: Needs and Pressures Steeming from Frameworks of Adult Competencies and Mathematics Curricula. In S.-J. Cho (Ed.), *Proceedings of the 12th International Congress on Mathematical Education (8 July – 15 July, 2012), COEX, Seoul, Korea* (pp. 1-7). Retrieved from http://www.icme12.org/upload/upfile2/tsg/2088.pdf.

Gal, I., & Garfield, J. (1997). Curricular Goals and Assessment Challenges in Statistics Education. In I. Gal & J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 1-13). Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter01.pdf.

Gigerenzer, G. (2004). *Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken [Calculated risk: How to know when numbers deceive you]*. Berlin: BvT Berliner Taschenbuch Verlag.

Gigerenzer, G. (2013). Risk savvy, how to make good decisions. New York: Penguin.

International Risk Governance Council (IRGC) (2005). *Risk Governance: Towards an integrative approach*. Genf: IRGC.

Kahneman, D. (2011). *Thinking fast and slow*. London: Allen Lane.

Kapadia, R (2008). Chance Encounters - 20 years later. Fundamental ideas in teaching probability at school level. *Proceedings of the 11th International Congress on Mathematical Education (6 July – 13 July, 2008), Monterrey, Mexico* (pp. 1-8). Retrieved from http://iase-web.org/documents/papers/icme11/ICME11_TSG13_24P_kapadia.pdf.

Kent, P., Pratt, D., Levinson, R., Yogui, C., & Kapadia, R. (2010). Teaching uncertainty and risk in mathematics and science. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics.* Lublijana, Slovenia: International Statistical Institute and International Association for Statistical Education. Retrieved from http://iase-web.org/documents/papers/icots8/ICOTS8_8C1_KENT.pdf.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.

Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse [Mathematical literacy: Development of the test and results]. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen*

Vergleich [PISA 2000. Students‘ basic competencies in international comparison] (pp. 139-190). Opladen: Leske+Budrich.

Kurz-Milcke, E. & Martignon, L. (2006). Educating children in stochastic modelling: Games with stochastic urns and colored tinker cubes. In A. Rossman, & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Salvador, Brazil: International Statistical Institute and International Association for Statistical Education. Retrieved from http://iase-web.org/documents/papers/icots7/C443.pdf.

Laskey, K. & Martignon, L. (2014). Comparing fast and frugal trees and Bayesian networks for risk assessment. In K. Makar (Ed.), *Proceedings of the Ninth International Conference on Teaching Statistics.* Flagstaff, USA: International Statistical Institute and International Association for Statistical Education. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_8I4_LASKEY.pdf.

Latten, S., Martignon, L., Monti, M., & Multmeier, J. (2011). Die Förderung erster Kompetenzen für den Umgang mit Risiken bereits in der Grundschule [Promoting risk literacy already in primary school]. *Stochastik in der Schule 31*(1), 17-25.

Mandl, H., Gruber, H. & Renkl, A. (1997). Situiertes Lernen in multimedialen Lernumgebungen [Situated learning in multimedia environment]. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia [Information and learning with multimedia]* (pp. 166-178). Weinheim: Psychologie Verlags Union.

Martignon, L. & Krauss, S. (2009). Hands-On Activities for Fourth Graders: A Tool Box for Decision-Making and Reckoning with Risk. *International Electronic Journal of Mathematics Education, 4*(3), 227-258.

Martignon, L. & Monti, M. (2010). Conditions for risk assessment as a topic for probabilistic education. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics.* Lublijana, Slovenia: International Statistical Institute and International Association for Statistical Education. Retrieved from http://iase-web.org/documents/papers/icots8/ICOTS8_8C2_MARTIGNON.pdf.

Moore, D. (1997). New pedagogy and new content. The case of statistics. *International Statistical Review, 65*, 123-165.

Riemer, W. (1991). *Stochastische Probleme aus elementarer Sicht [Stochastic problems from an elementary perspective]*. Mannheim: BI-Wissenschaftsverlag.

Savage, L. J. (1954). *The Foundations of Statistics*. New York, Wiley.

Sedlmeier, P. & Gigerenzer, G. (2001). Teaching bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*(3), 380-400.

Slovic, P. (2010). The psychology of risk. *Saúde e Sociedade, 19*(4), 731-747. Retrieved from http://www.scielo.br/pdf/sausoc/v19n4/02.pdf.

Spiegelhalter, D. (2011). Quantifying uncertainty. In L. Skinns, M. Scott, & T. Cox (Eds.), *Risk* (pp 17-33). Cambridge: CUP.

Spiegelhalter, D., & Gage, J. (2014). What can we learn from real-world communication of risk and uncertainty? In K. Makar (Ed.) *Proceedings of the Ninth International Conference on Teaching Statistics.* Flagstaff, USA: International Statistical Institute and International Association for Statistical Education. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_PL2_SPIEGELHALTER.pdf.

Taleb, N. N. (2007). *The black swan: The impact of highly improbable*. New York: Random House.

Till, C. (2014). Fostering Risk Literacy in Elementary School. *Mathematics Education, 9*(2), 83-96.

A Models and Modeling Approach to Risk and Uncertainty

Corey Brady
Northwestern University, USA
Richard Lesh
Indiana University, USA

Abstract: In this article we describe potential contributions of a Models and Modeling Perspective to research focused on learners' developing conceptions about uncertainty and variation. In particular, we show how a particular class of realistic problem-solving tasks can illuminate how learners develop models to identify, describe, and predict emergent patterns of regularity in the behavior of various types of systems and in the data these systems generate. We begin by situating current design work in this area within a larger project to investigate idea development in the domain of data modeling over extended (course-length) periods. We give design principles and examples for key components in our research framework, and we provide illustrative examples of these components and their interactions around the themes of distance and measurement that arise centrally in our materials. Next, we show how our approach can support advances in research on risk perception and on the development of ideas around risk assessment and management. Specifically, we identify three key facets of our approach and materials that make them good candidates for contributing to risk-oriented design research in education. Within each of these facets, we suggest research questions that could be addressed, and we provide examples and conjectures based on prior and ongoing work. In particular, we return to the ideas of distance explored in our examples and show connections with important questions in research on learners' perception and reasoning about risk.

Keywords: modeling, mathematization, problem-solving, affective dimensions of knowledge.

Introduction

In this article we describe potential contributions of a Models and Modeling Perspective to research focused on learners' developing conceptions about uncertainty and variation. In particular, we show how a particular class of realistic problem-solving tasks can illuminate how learners develop models to identify, describe, and predict emergent patterns of regularity in the behavior of various types of systems and in the data these systems generate. We begin by situating our current design work in this area within a larger project to investigate learners' ideas in the domain of data modeling, as they develop over extended (course-length) periods. Our design is centered on a type of activities known as Model-Eliciting Activities (MEAs), which engage learners in a deep form of modeling to construct solutions to real-world problems. The mathematical concepts uncovered in these core activities are then explored through Model Development Sequences (MDSs), where the classroom group of learners unpack and extend their collective ideas, connecting them with more formal mathematical constructs and investigating the reach of these new conceptual tools. We give design principles and examples for key components in our research framework, and we provide illustrative examples of these components and their interactions around the themes of distance and measurement, which arise in student solutions to MEAs and can be explored further through MDS activities.

After describing our framework, we show how our approach can support advances in research on risk perception and on the development of ideas around risk assessment and management. Specifically,

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 184-202

we identify three key facets of MEAs that make them good candidates for contributing to risk-oriented design research in education:

1. MEAs engage learners in optimization processes involving constraints and tradeoffs.

2. MEAs prompt learners to draw upon a wide range of problem-solving resources.

3. MEAs move beyond simplified problem settings that solicit the application of mathematical concepts and procedures that have been previously taught. Instead, by giving learners computational tools and pushing them to develop new mathematical constructions, MEAs are supportive environments for the kind of mathematical work that occurs in authentic settings outside of school.

Within each of these facets we suggest research questions that could be addressed through our approach, and we provide examples and conjectures based on our prior and ongoing work. To illustrate the importance of connections among ideas throughout a course-length engagement with modeling of this type and *across* MDS sequences, we suggest connections between the ideas of distance that were explored in our examples with important questions in research on learners' perception and reasoning about risk.

Theoretical Framework: Research in the Models and Modeling Perspective

In this section we briefly outline the theoretical perspective that underlies our design work. For over thirty years, researchers adopting a Models and Modeling Perspective (M&MP) in mathematics education (Lesh, 2003a; 2003b; Lesh & Doerr, 2003) have engaged in research to understand the development of mathematical ideas. A fundamental principle underlying this work has been that learners' ideas develop in coherent conceptual entities, called *models*, described by Lesh & Doerr (2003) as:

> …conceptual systems (consisting of elements, relations, operations, and rules governing interactions) that are expressed using external notation systems, and that are used to construct, describe, or explain the behaviors of other system(s)—perhaps so that the other system can be manipulated or predicted intelligently. (p. 10)

Under appropriate conditions, learners' models can be evoked and expressed in "thought-revealing artifacts." These artifacts can become objects for reflection and discussion by both individual learners and collaborative groups, and they also present rich data sources for researchers. In particular, when individuals and groups encounter problem situations with specifications that demand a model-rich response, their models are observed to grow through relatively rapid cycles of development toward solutions that satisfy these specifications. Models are thus powerful elements both for creating educational activities and for conducting research into learning.

From the Models and Modeling Perspective, Knowledge is seen as:
- Involved in perception and intuitions as well as action and conscious thought.
- Situated and shaped by context.
- Socially-shared and shaped by the community.
- Connected.
- Systemic, distributed, and emergent.
- Expressed in a variety of external media.
- Not simply logical / mathematical in nature.
- Often tacit.
- Initially piecemeal, undifferentiated, un-integrated and unstable.
- Continually developing along a variety of interacting dimensions.

Figure 1. Features of the M&MP conception of knowledge

As a program of research, M&MP research developed explicitly to investigate the following kinds of questions:

- •How can we characterize realistic problem-solving situations where solutions demand elementary-but-powerful mathematical constructs and conceptual systems?
- •What kinds of "mathematical thinking" are emphasized in such situations?
- •What does it mean to "understand" the most important of these ideas and abilities?
- •How do such competencies develop, and what can be done to facilitate their development?
- •How can we document and assess the most important (deeper, higher-order, or more powerful) conceptual achievements that are needed for full participation as citizens in increasingly complex societies and professions?
- •How can we identify students who have exceptional potential that is not adequately measured by standardized tests?

These questions are tightly linked to the M&MP's view on the nature of knowledge (see Fig. 1). Here, the M&MP builds on perspectives originating with Piaget and Vygotsky as well as with the American Pragmatists including Peirce, James, Meade, and the later Dewey, (*c.f.,* English et al, 2008; Lesh & Doerr, 2003).

Model-Eliciting Activities (MEAs)

Rooted in these perspectives and pursuing questions such as the ones listed above, M&MP research has sought to illuminate the nature of knowing and learning in authentic problem-solving settings. A key requirement of such settings is that they challenge learners to engage in original mathematical work (i.e., to produce mathematics constructions that are *new to them)*, rather than merely applying mathematics learned from an authoritative source. Iterative design work to create such learning environments has led to the development of a genre of materials and activities, known as Model-Eliciting Activities (MEAs). In MEAs, students are presented with authentic, real-world situations where they repeatedly express, test, and refine or revise their current ways of thinking as they endeavor to generate a structurally significant product—a *model*—comprising a conceptual structure for solving the given problem. These activities give students the opportunity to create, adapt, and extend scientific and mathematical models in interpreting, explaining, and predicting the behavior of real-world systems.

Originally designed as environments for research into what it means to "understand" important concepts in the K-12 mathematics curriculum, MEAs were first and foremost intended to provide documentation and evidence to illuminate the development of ideas in classroom groups. Thus, MEAs were designed in such a way that students would clearly recognize the need to develop specific constructs – without dictating *how* they would think about relevant mathematical objects, relationships, operations, patterns, and regularities. In general, this approach is inspired by the way engineers are given design "specs," which include brief descriptions of goals and available resources or constraints (such as time or money). As learning environments MEAs are also designed to optimize the chances that significant conceptual adaptations will occur during sufficiently brief periods of time so that the processes of conceptual change can be observed directly by researchers and teachers. Along with a range of particular MEA activities, the early M&MP community outlined six key design principles for MEAs to meet these goals (see, *e.g.,* Doerr & English, 2006; Lesh et. al., 2000; Hjalmarson & Lesh, 2007):

1. *Personal Meaningfulness.* Is the problem situation realistic, in the sense that a solution would be of genuine interest to a client? Is the problem space sufficiently open to ensure that different

groups of students are able to pursue diverse solution paths based in their own unique personal knowledge and experiences?

2. *Model Construction*. Does the problem truly require the new construction, modification, adaptation, or extension of a model in order to be solved? Does the problem engage with deep mathematical structures and regularities, rather than engaging mainly at the surface level?

3. *Self-Evaluation*. Are the problem's criteria sufficiently clear that student groups can judge for themselves the usefulness or adequacy of proposed solutions?

4. *Model Generalizability*. Do the models that are created in the activity apply only to the specific situation of the problem, or are they likely to be generalizable to a broad range of situations?

5. *Model Documentation*. Will student responses to the problem explicitly reveal their characteristic ways of thinking about the situation? Will they provide clear evidence about the mathematical objects and relations they have engaged with in solving the problem?

6. *Simplest Prototype*. Is the problem situation as simple as it can be, while still meeting the other design principles? Does the experience of the MEA "stick" with students so that they are able to use it as a lens for viewing future problems that feature similar mathematical structures?

An Example MEA: The Darts Problem

To illustrate the genre of the MEA, we provide the example of the Darts problem. This problem plays an important role in our work with data modeling, statistics, and probability, as, among other things, it challenges students to invent notions associated with centrality, spread, and distance from an expected distribution. Students are usually introduced to the Darts problem after having engaged with several other MEAs that press them to create operational definitions for key constructs such as "worker productivity" or "volleyball-playing ability," using data of various kinds to develop and apply quantitative measures of such constructs. The Darts problem pushes students to expand on this line of thinking, creating operational definitions of constructs for use in evaluating both performances (darts games) and performers (darts players).

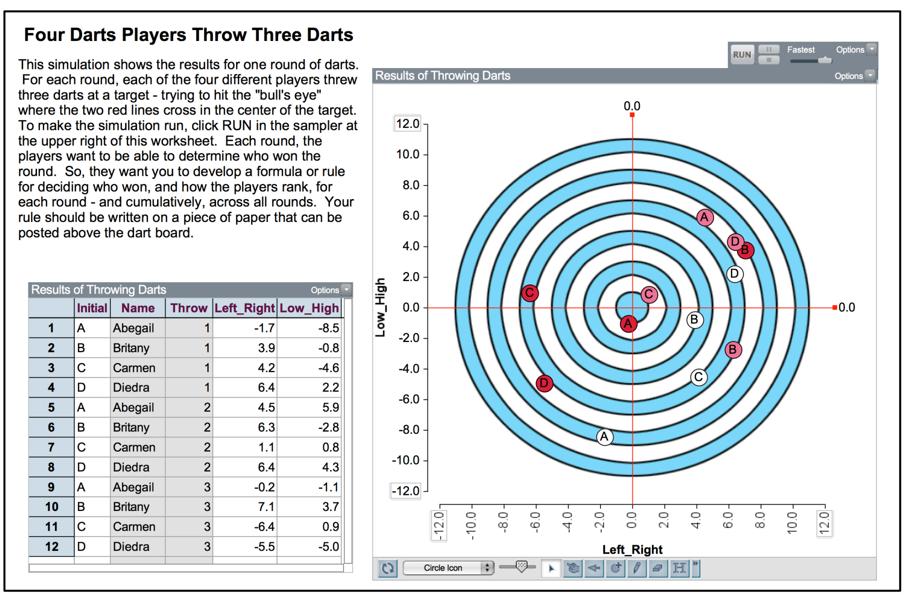


Figure 2a. Part one of the Darts MEA.

Darts, Part Two: ***Going for the Championship***
The local pub is now preparing for a season of competitive darts, and you need to provide useful information to the captain, Laura, to help her field a team. Beyond the four players from Part One, a group of "regulars" at the pub will be trying out. Describe how results from several rounds of darts can be used to create a rating or description of players' darts throwing skills. Your procedure should help Laura to choose a team based on the tryouts.

Figure 2b. Part two of the Darts MEA.

After breaking into groups, the students are presented with the two-part problem statement (Figures 2a and 2b, above). The Darts problem is somewhat unusual among MEAs in being itself embodied in software form, using a dynamic data exploration tool (here, TinkerPlots). In contrast, most MEAs are paper-based, with dynamic mathematics software being more heavily used in follow-on activities. The basic structure of students' engagement with the Darts problem, however, is typical of MEAs. Student groups engage in cycles of thinking, characterized by identifying possible interpretation schemes for the problem context, testing these schemes, and revising or adapting them to accommodate new ides or to address shortcomings. Through these cycles, their thinking evolves rapidly towards increasingly effective approaches to the problem. For instance, in the first 10 minutes of work, groups may focus on one or another feature of the problem (e.g., attending to one dart of the three thrown per player, or even, when attending to the coordinate values for a single throw, attending only to the greater of the two). As they grapple with and discuss the first part of the problem, however, groups generally move toward a definition of distance-from-the-bullseye. (This is often expressed either as Euclidean distance, as its square (adopted for convenience in calculation), as single-coordinate differences, or as a "taxicab" measure obtained by summing these coordinate distances (as absolute values of coordinate differences.))

While the idea of distance is sufficient for the task of ranking individual throws according to quality (with the rule that a lower distance corresponds to a better throw), it does not solve the problem of scoring a round. Recognizing this, some groups turn to a notion of an average (either based on their distance calculations or re-invented independently), while other groups attend to the rings drawn on the board, associating scores with each ring by drawing on intuitions associated with fair-rewards. To support this second strategy, groups may tap into ideas about probability and area, beginning to reflect on the likelihood that a "random throw" will land inside different target circles.

In tackling part 2 of the Darts problem, the groups' focus generally moves from scoring *throws* and *rounds* to evaluating *players*. Here, groups begin to generate candidate attributes for players such as "consistency" "precision" and "accuracy." In many cases, learners realize that their groupmates have slightly different attributes in mind for these terms. Many groups find that (at least) two are at work—one corresponding roughly to an idea of "centrality" and another corresponding to ideas of "spread." Group discourse also begins to reveal questions about the nature of simulations and distributions, as students question whether the computationally simulated players "really" have different skill levels or attributes, and if so, how they might detect such differences with repeated trials of the simulation. Finally, additional concepts emerge as groups converge on a solution and begin to draft their letter to Laura, the "client." For instance, they may reflect on which of the two attributes might be more "teachable" to players over the season, or they might become concerned about how to reduce the logistical complexity for judges attempting to choose between candidate players in noisy pub settings.

In this way, students iteratively develop solutions to the problem in the time allotted—usually 60 minutes for this MEA. In some implementations, the class then gathers together for a structured "poster session" event. Here, one member in each 3-person group hosts a poster presentation showing the results of their group. The other two students use a Quality Assurance Guide to assess the quality of the results produced by other groups in the class. These forms are submitted to the teacher and contribute to assessment in various ways, providing evidence for the modeling and conceptual achievements of both individuals and groups.

Model Development Sequences (MDSs)

Moving beyond individual MEAs, recent M&MP research has also investigated ways in which MEAs can be integrated within larger instructional sequences: Model Development Sequences, or "MDSs" (Lesh et. al., 2003). MDSs offer classroom groups opportunities to unpack, analyze, and extend the models they have produced in MEAs, as well as to connect their ideas with formal constructs and conventional terminology. This unpacking work helps to ensure the lasting retention of concepts at the level of generality required to apply them flexibly in novel situations. MDS activities also set the stage for the critical connection between conceptual development (the centerpiece and focus of MEAs), and procedural knowledge that is required for students to achieve well-rounded competence in any subject area. In our examples here, we will highlight how the key theme of *distance* that emerges in students' MEA work can be developed in MDS activities.

Within an MDS, *reflection tools* support students in stepping back from their modeling processes and reviewing this work as critical observers of both individual and group modeling behavior. In the design of our course materials, we consider these tasks core to the learning process. In general, M&MP research expects that when students interpret situations mathematically, the interpretation systems they engage are not purely logical or analytical in nature. Rather, they also involve attitudes, values, beliefs, dispositions, and metacognitive processes. Moreover, the M&MP does not treat group roles or group functioning as if these were fixed student attributes that determined their behaviors. Instead, students are expected to develop a suite of problem-solving *personae* that they learn to apply purposively as the situation demands.

In *product classification* and *toolkit inventory* activities, students continue the work of abstraction, identifying links among their solutions to different MEAs and between these solutions and the "big ideas" of the course. *Model exploration activities* (MXAs) provide a model-rich environment for introducing more conventional terminology, concepts, and skills, which students need in order to formulate sophisticated models and present them to a mathematical community. These may use a combination of pointed YouTube videos and interactive simulations in dynamic mathematics software. (In our work to construct a course-sized repository of materials to fuel MDSs there are to date approximately 50 of these YouTube videos with accompanying simulations in dynamic geometry and dynamic statistics software. These are currently collected under the ProfRLesh channel.) Finally, *model adaptation activities* (MAAs) allow students to transfer ideas and techniques developed in MEAs to situations calling for similar performances. These MAA activities also provide smaller-timescale modeling scenarios that exercise concepts students have explored in other components of the MDS. They may be pursued individually or in small groups, depending on the nature of the task and the teacher's instructional or assessment goals.

Example MDS Activities Associated with the Darts MEA

Model Exploration Activity (MXA)

As mentioned above, most student groups develop one of several distance formulas in order to evaluate the quality of dart *throws* in Part 1 of the Darts MEA, and most also operationally define an attribute of a darts *round* that is analogous to a construct expressing the *center* of a dataset. They then examine the variation of this attribute over many rounds to describe an attribute of the *player*. A MXA that helps to unpack and extend this work is shown in Figure 3 below, implemented using the NetLogo software (Wilensky, 1999). Here the NetLogo environment produces visualizations of "level curves" of the sums of distances to chosen points. Students can add, remove, and rearrange these points, and they can select among various definitions of distance. Screens 3a, 3b, and 3c show the different results obtained by using three points and varying the distance definition (3a uses the square of Euclidean distance, 3b the Euclidean distance, and 3c the "taxicab" or "Manhattan" distance). Screen 3d shows the effects of adding points (there are five) and zooming out in scale. The numerical values for distances and the coloring update in real time as points are moved, giving students a multidimensional sense of the way different distance functions operate. Moreover, because NetLogo is a "glass box" environment, it is a relatively simple matter to modify distance formulae or add new ones, change the coloring visualization, and so forth.

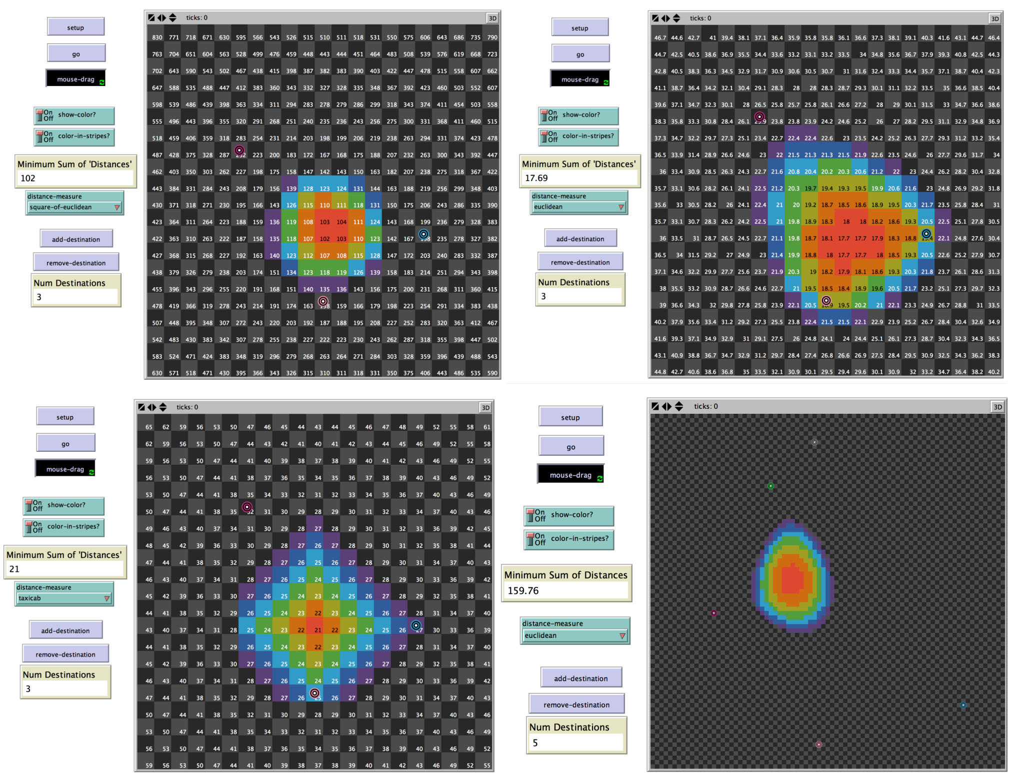


Figure 3. A dynamic environment for visually exploring the effects of different distance formulae. Fig 3a (upper left) Square of Euclidean distance; Fig 3b (upper right) Euclidean distance; Fig 3c (lower left) taxicab distance; Fig 3d (lower right) Euclidean distance, with five points, zoomed out.

Though learners are given freedom to explore in this environment, developing intuitions and a feel for how distance measures are affected by changes in the data, there are also several directions that teachers may wish students to explore and discuss more systematically. One such direction involves investigating what happens when all data points are collinear. In this case, the points can be interpreted as scalar values along the shared line, and the different distance-minimizing centers can be connected with the familiar notions of the mean and median of a dataset. Another involves studying the effects of *parity* on the different distance definitions – how it sometimes matters whether there are an odd or even number of points in the dataset. Finally, one can explore questions of weighting by making two or more points coincident and investigating the impact on the various centers. In general, moving to the two-dimensional setting helps to ground many of the definitions that have been learned with one-dimensional (scalar) data in a broader context.

Toolkit Inventory Activity

While completing the Darts MEA and reviewing their peers' solutions, students may collectively identify such modeling toolkit entries as "simulating many trials" or "eliminating outliers." In terms of the "big idea" of distance, questions about how to identify or define outliers arise naturally from learners' impulses to offer darts players "do-overs" under some conditions or to discard certain throws as not being "accurate representations" of the players' skill level. Indeed, a somewhat unusual but by no means rare strategy for scoring a round is to discard the best and worst throws, basing the score solely on the "median" throw. While this strategy may drop out in favor of other approaches, the idea of discarding data may reappear in different guises, and the question of its fairness can draw out debate on a variety of important topics, including the nature and identification of outliers. This can be a very rich topic for classroom discussion, as it integrates a variety of key ideas including notions of representativeness, rarity, variation between and within individuals, central tendency, and distance.

Model Adaptation Activity (MAA)

Again building on the idea of generalizing distance from a landmark outcome (the bullseye in the Darts MEA), students can be asked to consider other simulations in which an outcome is a complex event, as, in the example shown below, 30 throws of a six-sided die. Though this situation is in some senses quite different from the Darts problem, students find that powerful ideas from that setting can guide their work in this new context.

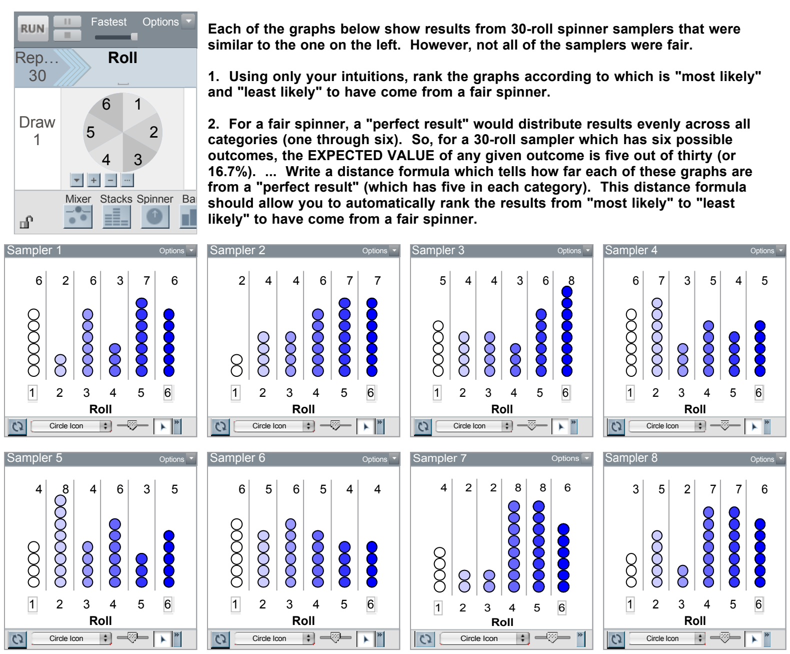


Figure 4. Model Adaptation Activity (MAA). Students generalize the idea of an invented distance from the Darts MEA to measure the distance of these distributions from an "expected" distribution.

Here, students are first asked to rank the outcomes intuitively according to their sense of which is more likely to occur. Discussions reveal some agreement (usually at the extremes of "likely" and "unlikely" distributions), with other points of dispute (usually among the middle-likelihood distributions). On one hand, the agreement helps to confirm the intuitive nature of applying the idea of a distance and the construct of an expected value in this setting. On the other hand, the disputes drive the class toward establishing an explicit, quantitative definition of this distance to settle arguments. After developing and coming to agreement on a distance concept and working through the process of operationalizing it as a formula, students can express that formula computationally in the TinkerPlots environment and run a large number of simulated trials of fair dice. Representing these trials according to their distance measure (or, if there are several candidate distance measures, according to each of these), students estimate the probability that sample's distribution will be a given distance, or further, from the expected distribution. (See Figure 5, below.) That is, students can observe that in, say, 10,000 trials using a fair spinner, only about 100 trials yielded a distance measure of greater than a threshold value. This would provide an empirical estimate of the *probability* of meeting or exceeding this threshold (of approximately .01).

Furthermore, the classroom group can use this simulation approach to reflect back on the distance measures themselves. For instance, they can confirm whether their measure does in fact correlate to likelihood (i.e., whether larger distances from the expected value are (uniformly) rarer). They can also make judgments about whether one definition or another does a better job of "separating" outcomes in different ranges. Distance measures can then more formally and confidently be connected with the probability of being that far or further from the expected value (essentially a one-tailed *p*-value). Finally, students can either learn that the procedure they have invented is essentially the *chi-square* measure, or they can compare their approach to that standard test. Importantly, this experience of the chi-square concept does not require a blind appeal to the authority of statistical tables or results from mathematical analysis beyond the level of middle- or high-school students, and it appropriately grounds such tools (when they are encountered) in intuitive notions of distance and expectation.

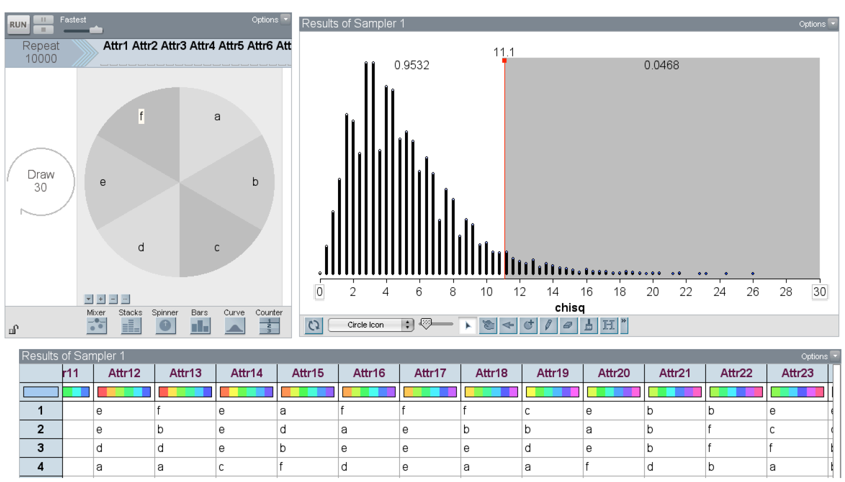


Figure 5. TinkerPlots simulation of 10,000 trials of 30 samples, evaluated according to the class's chosen definition of "distance from the expected distribution"

Three Facets of MEAs that Make them Propitious Environments for Research on Risk

Having now given an account of M&MP research in general and our approach to MEAs and MDSs in particular, we are in a position to discuss the application of this design research methodology to study questions associated with perceptions and reasoning about risk. In particular, we describe three facets of MEA research that resonate with risk-related education research are as follows:

1. They engage learners in optimization processes involving constraints and tradeoffs. MEAs present learners with problem settings that require them to optimize solutions and processes, working within constraints, making trade-offs, and accounting for second-order effects. Risk can play a key role in such situations, introducing the need to balance prospective gains against potential losses. Learners' invented strategies can provide a rich source of information about their ideas and attitudes about both.

2. They prompt learners to draw upon a wide range of problem-solving resources. MEAs promote reasoning and the use of knowledge resources in a broad sense, including not only logical/analytical thinking but also affective/intuitive thinking. As such, the view of modeling that emerges from such activities includes the feelings, attitudes, values, and beliefs of learners. Recent research has become increasingly tuned to the role of feelings and intuitions both in conditioning perceptions of risk and in generating the human meaning behind abstract statistics and probabilities; thus, these dimensions of problem solving may be of particular interest (see, e.g. Slovic, 2000, 2010; Borovcnik, 2011).

3. They move beyond simplified problem settings that solicit the application of mathematical concepts and procedures that have been previously taught. MEAs involve learners in the interpretation of phenomena and situations that are not pre-classified as examples of constructs taught in a particular topic or textbook area. Thus, they can offer excellent opportunities for exploring the development of ideas associated with a variety of conceptions of uncertainty that are interconnected in real-world settings but that are taught and learned in different domains and at different points in school curricula. Because MEAs involve learners in authentic creation of mathematical models, they offer designers the opportunity to engage learners with phenomena that may evoke very different conceptions of indeterminacy, ranging from randomness and error on one hand to complexity and emergent phenomena on the other. Perceiving and managing risk may take on very different aspects in these different settings.

We deal in more depth with each of these facets in the sections below. For each facet, we discuss potential contributions of M&MP research and we outline research questions and conjectures that can be pursued by the design research approach we have outlined above.

Facet 1. MEAs engage learners in optimization processes involving constraints and tradeoffs

In MEAs learners are presented with challenges that require them to optimize processes or outcomes, work within constraints, make trade-offs, and account for second-order effects. Many problems involving risk are similar in the sense that they require balancing the opportunity for some kind of gain with the potential for some kind of loss. Such modeling situations are conceptually rich in part because they demand a deep and flexible understanding of the systems that are involved. To understand these features, learners must envision the *parameter space* of a system—the range of its behaviors as its key parameters change. Essentially, this replaces 'snapshot' evaluations of a system's behavior with a sense of multi-dimensional covariation. Moreover, when the consistency of a system's behavior and/or the measurement of that behavior are themselves problematized, multiple notions of uncertainty and variation are introduced and must be coordinated. MEAs can place students in situations that require them to develop models of systems under these kinds of conditions; as a result, they offer extremely rich environments to study learners' ways of thinking and the broader development of ideas about these topics

Risk- and uncertainty-related research questions that could be pursued via MEA designs include questions like the following:

•When faced with realistic problem settings, what kinds of uncertainty and variation are most salient to learners and enter most strongly into their calculations and their proposed solutions?

•Do learners develop different coping or mitigation strategies for different categories of variability in their data?

•How does the frame of recommending a course of action for someone else (e.g., the "client" of the MEA) affect learners' assessments of risk and the weight they assign to possible outcomes?

Finally, in a course-length engagement with MEAs and MDSs, there are interactions with big ideas and constructs developed elsewhere. Taking as an example the idea of *distance*, as discussed above, we might ask whether and how learners draw on such multi-dimensional notions of distance as resources for expressing and operationalizing "riskiness." Similarly, we could attend to learners' perceptions of weaknesses or limitations in using the distance construct as a definition of riskiness, and we could use these responses to illuminate their reasoning processes.

Facet 2. MEAs prompt learners to draw upon a wide range of problem-solving resources

Research in modeling indicates that learners develop *problem-solving personae* over an extended period of engaging with problems like those found in MEAs (Hamilton, Lesh, Lester, & Yoon, 2007). Investigations using the Reflection Tools that are part of MDS designs described above increasingly suggest that while these problem-solving personae may have a logical or technical core, they also involve "soft" aspects of knowledge, including attitudes, feelings, and beliefs (both about oneself and the domain). Here, we find a strong potential overlap with recent studies on the psychology of risk perception, which emphasize the impact of non-logical processes on learners' conceptions of risks and probabilities. Particularly interesting in this regard is evidence supporting *dual process* theories of probabilistic thinking and responses to risk (e.g., Kahneman, Slovic, & Tversky, 1982; Kahneman & Frederick, 2002). Such research posits the existence of "heuristics" that support intuitive, "System 1" responses to real-world situations. More generally, the role of feelings in human responses to risk and uncertainty is increasingly important in this domain, as led by work of Slovic and colleagues (e.g., Slovic, 2000, 2010). Their research has established an "affect heuristic" (Finucane et al, 2000; Slovic et al, 2002), by which "feelings serve as an important cue for risk/benefit judgments and decisions." (Slovic, 2010, xxi).

In spite of evidence that people do in fact give great weight to their intuitive, experiential sense of risk, as compared with more deliberative and analytical sense-making processes, Slovic and colleagues do *not* argue in favor of extirpating the intuitive in favor of the analytical. On the contrary, their research suggests that intuitive and experiential factors play a key role in grasping the *meaning* of probabilistic situations, so as to weigh risks appropriately (Slovic, 2002). That is, this line of research suggests "we, as a species, think best when we allow numbers and narratives, abstract information and experiential discourse, to interact, to work together" (Slovic, 2010, p. 79). At the same time, there are many situations in which humans can be misled by their feelings. A reasonable approach to this dilemma would seem to involve supporting learners in building *both* a better toolkit for analytical thought *and* more effective ways of associating or formulating stories, scenarios, and images for data, as resources for affective responses (see also Borovcnik, 2011).

Of course, understanding the affect heuristic and its relation to other dynamics of risk perception and reasoning is an ongoing challenge. Here, the design of MEAs can offer researchers a range of tools for getting at related questions. For instance:

•How does the social setting of a small problem solving group change the nature of risk perception in its members? Does in-the-moment social discourse provide resources that can be used by the group to blend System 1 and System 2 responses?

•How durable is affect-based risk perception? Does it leave traces in the solution strategies formulated by groups and/or in their communication strategy for explaining their solutions to the "client" of the MEA?

Finally, connecting again with the distance theme, when the likelihoods of threats and risks (or, benefits and rewards) are described with reference to distance, does this promote new and/or unanticipated ways of thinking about contingency? A spatial metaphor may, for example, suggest to learners notions of intervening regions or "buffers" affected by and depleting the threat or benefit.

Researching Facet 2: Early design of an MEA and MDS on Probability and Risk Perceptions

We have not yet incorporated activities specifically focused on risk perception and reasoning with our existing materials on data modeling and statistics. However, we have begun the design process of creating an MEA and MDS to study related issues, and we describe some of our conjectures here. The process of designing MEAs is itself an iterative modeling process, and so we expect that our thinking in this area will develop significantly as we proceed.

A preliminary statement of a prototype-MEA, the Charity Benefit problem, is shown below:

Charity Casino Night Benefit

In this activity, your task is to design two different chance-based gambling booths for a charity benefit event. At charity benefits, the attendees are in principle happy for the "house" to win more than the players, but they still want the games to offer them a reasonable chance of winning. In fact, they may be more likely to play the games where they think they are more likely to win.

The organizers of the benefit expect approximately 500 people to attend the event. As you design your two booths, be sure to explain to the benefit organizers:

- how each game is played,
- how much the organizers should charge people to play each game, and
- how much each game should pay out to winners.

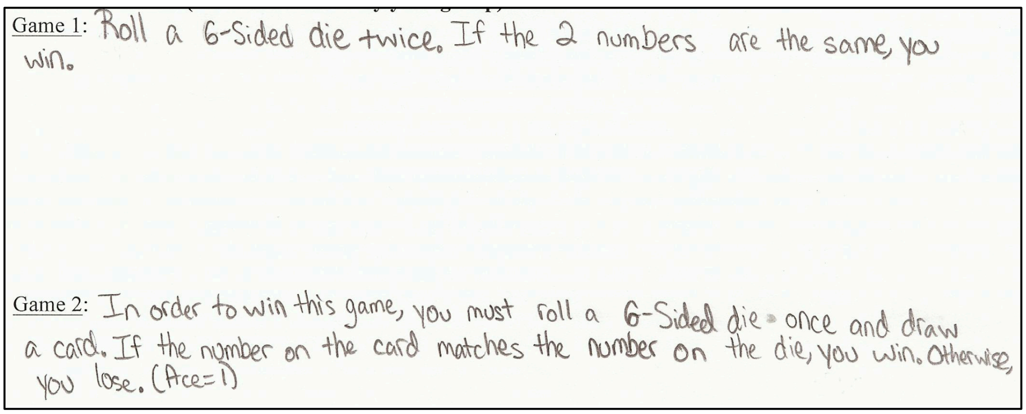
NOTE: In writing your game's rules and describing a player's winnings, be sure to clarify how the player's stake (what they paid to play) is treated when they win. For example, if the game costs $1 to play and winners receive $2, does this mean that they receive $2 *plus* the $1 they staked, or $2 *including* the $1 they staked?

Write a letter to the benefit organizers including the description of your game and any guidance you can give them about what they can expect to earn from your two booths in the one-night charity benefit event.

Figure 6. The Charity Casino Night problem.

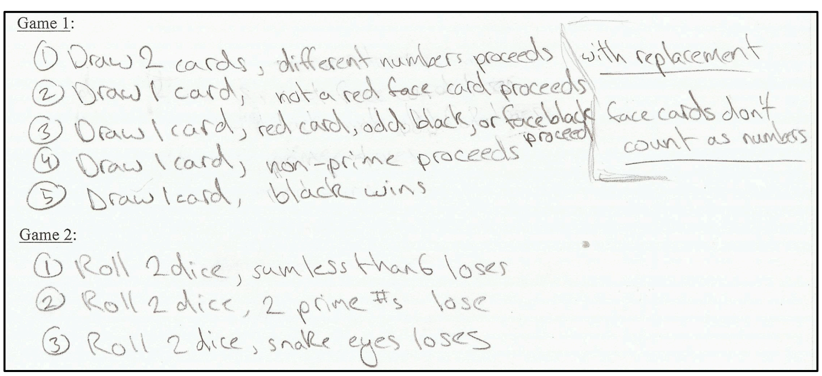
In building toward this problem, we explored several questions with small groups of 7th-9th grade learners from a weekend enrichment program. The first question was whether the students would find it interesting to invent chance-based gambles. We found this to be both an engaging and a revealing activity, particularly when the task was phrased in terms of creating gambles that would fool prospective players about the likelihood of winning. Our most recent activity to test the viability of this type of design task was as follows: Each group worked independently to design a pair of chance-based gambles. When all groups had completed their designs, one group was chosen to present. An opposing group (the "challengers") was offered their choice of the two, and the students themselves (the "authors") were then forced to take the other gamble. The challengers and the authors then played their gambles until one side won. Following the challenge, the authors explained their design, with an emphasis on demonstrating which side "should have won" based on the relative probabilities of winning the two gambles.

In the students' designs, "analytical" strategies (such as creating gambles whose probabilities were difficult to compute) were deeply mixed with "psychological" or "rhetorical" strategies, in which the authors attempted to manipulate their opponents' perceptions of relative probabilities through various means. Figures 7-9, below, show three groups' challenges. In Figure 7, the group hoped that the phrasing of Game 1 would trigger their opponents to assess the probability of winning as 1/36 rather than 1/6. In Figure 8, the group used contrasting descriptions (descriptions of winning in Game 1 versus descriptions of losing in Game 2) to attempt to manipulate their imagined opponents' assessment of probabilities. In Figure 9, the group was hoping their opponents would focus of the number of winning balls, instead of the proportion of winning outcomes to total outcomes. (In fact, the students' calculation of the probability for game 2 was incorrect, so Game 1 actually had slightly more favorable odds.)


Game 1: Roll a 6-Sided die twice. If the 2 numbers are the same, you win.

Game 2: In order to win this game, you must roll a 6-Sided die once and draw a card. If the number on the card matches the number on the die, you win. Otherwise, you lose. (Ace=1)

Figure 7. Students are attempting to cue the idea of a 1/36 probability for double sixes in Game 1.


Game 1:

1 Draw 2 cards, different numbers proceeds | with replacement
2 Draw 1 card, not a red face card proceeds
3 Draw 1 card, red card, odd black, or face black proceed | face cards don't count as numbers
4 Draw 1 card, non-prime proceeds
5 Draw 1 card, black wins

Game 2:

1 Roll 2 dice, sum less than 6 loses
2 Roll 2 dice, 2 prime #s lose
3 Roll 2 dice, snake eyes loses

Figure 8. By describing winning conditions in Game 1 and losing conditions in Game 2, this group hoped to lure opponents into choosing the less-favorable gamble.

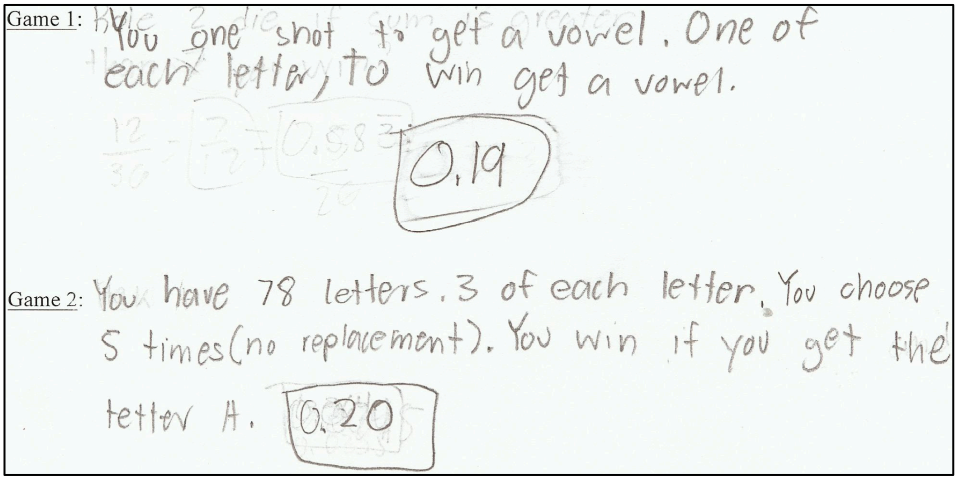


Figure 9. There are more “winning balls” in Game 1, but more chances to draw in Game 2.

This early experiment has suggested the potential of designs in which students attempt to manipulate the presentation of a gamble to make it more likely for another person to mis-perceive the odds. To capitalize on this feature in the Charity Benefit MEA, we plan in our next iteration to convert the “poster session” of the MEA into a simulation of the charity event itself. To do this, we will use a networked participatory simulation or “PartSim”(Colella, 2000; Colella, Borovoy, & Resnick, 1998; Wilensky & Stroup, 1999a; Klopfer, Yoon & Perry, 2005), implemented using the HubNet module (Wilenksy & Stroup, 1999b) of NetLogo. In this simulation one group member will “man” the group’s booths while their groupmates circulate in the virtual charity casino, spending an allotment of tokens on the available games.

In other work related to social policy and risk of flooding (Brady et. al, in preparation), we have seen the power of such PartSims to provide learners with a “feeling of risk” (Slovic, 2010), even when the associated computer visualizations are quite simple and do not provide an immersive “virtual reality” experience. In the Charity Event PartSim, we aim to provide two kinds of risk sensation: one, for the two roving players, as they make selections similar to those that gamblers at the event would do; and a second, for the booth-manning team member, who can monitor many more instances of game play and watch the winnings for the booths, compared to the expectations they have stated in their designs. Students will circulate among roles, experiencing both. If our flood modeling experience is a good indicator, students in both roles will feel substantial and “realistic” levels of tension about the results, even though no actual currency is at stake in the gains or losses of the PartSim.

Patterns that emerge in students’ interactions with the PartSim will also create realistic feedback to the groups about the relative attractiveness of their games. Games with more attractive winning propositions may be played more often, and this will give them a greater opportunity to return revenue as expected by the theoretical probabilities that underlie their gambles. On the other hand, the “casino” environment will also introduce complications and complexities that should make the analysis of the experience and the data that results into rich contexts for reflection and discussion. For instance, games that happen to return atypical results in their early plays or games that return variable winnings may experience changes in popularity that the group can reason about in terms of such concepts as information cascades on the one hand or the attractiveness of certain kinds of winnings profiles on the other. In any case, the tabulated results on the net revenues earned for the charity will be a dataset worthy of group reflection, as will be the various experiences and perceptions of the learners-as-players.

Facet 3. MEAs move beyond simplified problem settings that solicit the application of mathematical concepts and procedures that have been previously taught.

M&MP research argues that models are *conceptual systems* that support and organize perception itself. In a wide range of settings, when we engage with the world our models enable us to *see* phenomena in particular ways: they make selected attributes of situations salient to us, and they orient us toward potential actions and judgments. Our *mathematical* models do this work in ways that emphasize structural features of situations. Among other things, this means that in realistic problem settings, experts distinguish themselves from non-experts not only by what they *do* but also by what they *perceive*.

This perspective also implies that authentic modeling is a fundamentally *interpretive* process. MEAs, like real-world problem settings, are not framed as occasions to apply a particular procedure or construct. Instead, they admit a variety of possible approaches, and most adequate solutions are characterized by bringing together ideas from different subject areas. In general, a modeler's entire store of prior knowledge and experiences acts as an interpretation system through which she or he is able to make sense of new phenomena. And in the cyclic modeling process of MEAs, learners build upon their past experiences and prior knowledge to develop new ways of seeing problem situations. Again, this is one of the primary reasons why solutions to MEAs often incorporate ideas from across the experience base of problem solvers. They may draw on ideas and techniques from multiple domains or textbook topic areas, and they may merge logical analyses and calculations with value-based judgments, feelings, and beliefs.

We argue that research is needed to investigate fundamental questions about students' reasoning and what it means to understand core ideas in probability, uncertainty, and risk. Given this, we believe it is essential to study learners' modeling processes in settings that feature open and authentic inquiry in the sense descried above. The openness and authenticity of the challenge is critical to the value of modeling activities and can be overlooked in discussions of modeling. For instance, in spite of an encouraging emphasis on "modeling" in the Common Core State Standards in Mathematics (CCSSM, 2010), it is unfortunately possible to see the description of modeling there as describing mere *applications* of already-learned mathematical constructs to simplified real-world situations. As an example, in the CCSSM practice standard, "Model with Mathematics" we read:

> Mathematically proficient students can *apply the mathematics they know* to solve problems arising in everyday life, society, and the workplace. In early grades, this might be as simple as *writing an addition equation to describe a situation*. In middle grades, a student might *apply proportional reasoning to plan a school event* or analyze a problem in the community. By high school, a student might *use geometry to solve a design problem* or *use a function to describe how one quantity of interest depends on another*. Mathematically proficient students who can apply what they know are comfortable making assumptions and approximations to simplify a complicated situation, realizing that these may need revision later. They are able to identify important quantities in a practical situation and map their relationships using such tools as diagrams, two-way tables, graphs, flowcharts and formulas. They can analyze those relationships mathematically to draw conclusions. They routinely interpret their mathematical results in the context of the situation and reflect on whether the results make sense, possibly improving the model if it has not served its purpose. (CCSSM, 2010, emphasis [with italics] added)

The distinction between modeling and application is essential, both to research and to teaching. When problems are offered to learners as mere applications of particular mathematical principles or concepts, students do not experience authentic, fundamental challenges of interpretation. In contrast, when learners are presented with realistically complex problems (before instruction), they may engage in modeling as we have described it: a fully interpretive struggle. Student thinking in these contexts can thus illuminate connections and relations between topics normally treated as separate. We argue that this is particularly true with probability and uncertainty, and we believe that this is one of the reasons to expect that MEA-based research can be illuminating in this area.

In fact, there are dimensions of research in risk and uncertainty that can be specifically illuminated by this form of learning activity. In closing this section, we outline one such topic: the distinction between sources and types of variation. In conducting authentic inquiry or making sense of novel situations, modelers often engage with various forms of uncertainty and may interpret them as arising from a variety of causes or mechanisms—for example, as due to *error*, to *chance*, or to *complexity*. Their tacit or explicit classification of these sources of uncertainty and variation may frame their modeling activity to highlight dimensions of the phenomena they describe, for example focusing on issues of measurement, on the action of unpredictable or stochastic systems, or on the operation of systems involving feedback loops, emergence, or sensitive dependence on contextual conditions.

Importantly, many rich or multi-layered phenomena admit more than one valid interpretation. Moreover, even when one source or type of variation is dominant, the quantitative data collected from such settings themselves do not always offer distinctive cues to guide correct classification. For instance, the distributions of Figures 10 and 11, below, arise from situations that are more or less clearly dominated by issues of measurement (10a), probability (10b), and complexity (11). Yet the data distributions themselves are all visually similar and do not betray the types of mechanisms that generated them. As with "distribution," so too with other core ideas involved in modeling risk and uncertainty, which also cut across these domains (e.g., variation, independence, sample space, or expected value).

Given that the data themselves may not determine the "correct" model or interpretation of the source of uncertainty in the situation, students have the opportunity to engage in authentic forms of argumentation, negotiation and inquiry. In such settings, students' different ways of thinking about phenomena and the different resources they draw on for interpretation can offer researchers insights about possible connections among ideas.

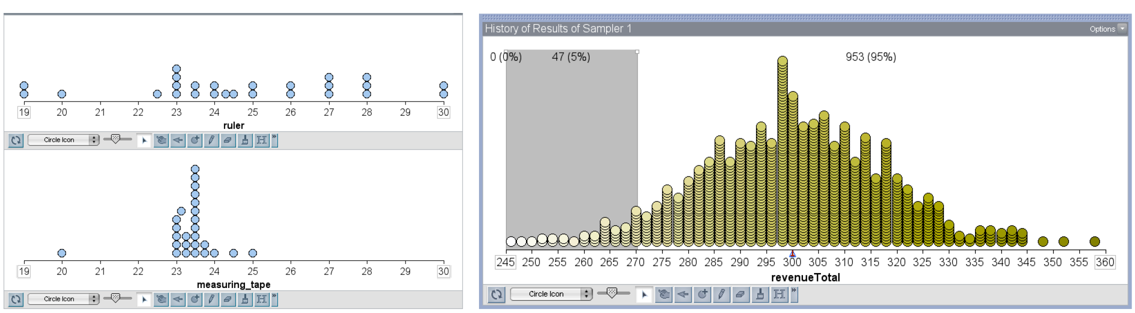


Figure 10. Measurement (left). Distributions in measures of the circumference of the teacher's head using a ruler and a measuring tape (after Konold & Lehrer, 2008). Probability and sampling (right). Revenue outcomes from one booth in 1,000 simulated charity events in the extended gamble activity.

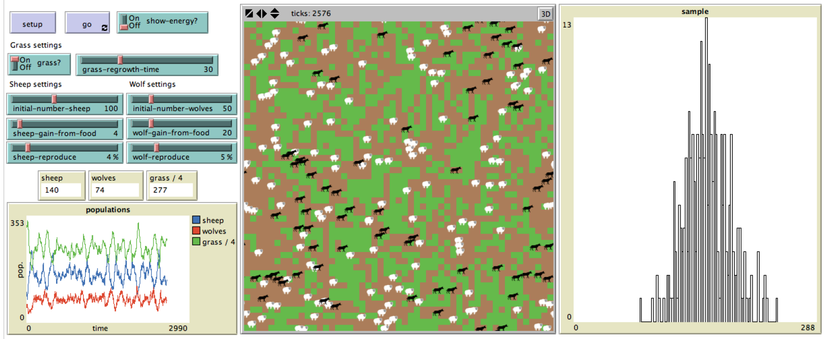


Figure 11. Complexity. Periodic sampling of the sheep population in a NetLogo simulated ecosystem.

Furthermore, in the course of developing solutions to modeling problems, learners are forced to take a range of *stances* toward the uncertainty they encounter in phenomena, depending on their immediate goal and focus. These stances include:

•*Conceptualizing the situation.* Here, the modeler invokes notions of uncertainty, chance, or complexity in making sense of the primary phenomenon (i.e., answering the question, "What kind of thing is it that we are describing?").

•*Gaining data or information about the situation.* In making or analyzing measurements of any phenomena, modelers grapple with error and imprecision. Further some forms of measurement may come to seem epistemologically problematic (i.e., these actions introduce a kind of "Heisenberg" principle, the urgency of which is heightened by the framing in terms of risk).

•*Describing and interpreting a collection of data about the situation.* In understanding situations of uncertainty, we engage such concepts as our "expectation," or our assessments of "best case and worst case scenarios." Different domains offer different tools for addressing these questions, even if most involve some form of statistical thinking. We have shown above how distance can act as such a tool, one that derives from students' experiences of geometry and physical measurement.

•*Combining and weighing these perspectives to formulate a strategy or decision in response to the situation at hand.* Returning to ideas of variation, managing risk may primarily involve *controlling* and *minimizing* variation, or *describing its cause* and *estimating its consequences* on some other process or decision. In such cases, substantially *different* approaches and lines of inquiry may suggest themselves depending on whether the modeler views the variation in the system as arising from error (e.g., "noise,"), from inherent randomness, or from complexity (e.g., "feedback" or "emergence").

In MEAs, as learners assume any of these stances toward the phenomena they are modeling, or as they shift among them, their work offers us different perspectives on their emerging thinking. Moreover, these stances toward a problem do not necessarily form a sequence: learners may cycle through them multiple times, with the thinking from each stance deeply affecting the others.

Conclusion

We have a great deal to learn about idea development in authentic problem settings where probabilistic reasoning, uncertainty, and risk are foregrounded. In this article we have argued that Model-Eliciting Activities and Model Development Sequences can offer promising settings for investigating key questions in this area. In closing, we note that in this area as in many others, learners' knowledge is constituted as much by *connections* forged between big ideas in the domain and between these ideas and prototype situations as by an "intrinsic" understanding of these big ideas in isolation. The process of learning may therefore be expected to be multi-dimensional and non-linear. Thus, research into idea development in this area needs to identify both (a) local operational definitions of what it means for students to have learned big ideas in the domain, and (b) longer-timescale accounts of students' growing appreciation of the significance and interrelatedness of these big ideas. We hope that the focus on learning processes across multiple MEAs and MDS units can contribute insights to this ongoing effort.

References

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.) *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (71-83). Springer: Netherlands.

Brady, C., Nay, J., Sengupta, P., Gilligan, J. & Camp, J. (in preparation) A Participatory Simulation and Agent-Based Model of Flooding.

Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences, 9*(4), 471-500.

Colella, V., Borovoy, R., & Resnick, M. (1998). *Participatory simulations: Using computational objects to learn about dynamic systems.* Paper presented at the Computer Human Interface (CHI) 1998 Conference, Los Angeles, CA.

Doerr, H. M., & English, L. D. (2006). Middle Grade Teachers' Learning through Students' Engagement with Modeling Tasks. *Journal of Mathematics Teacher Education*, *9*(1), 5-32.

English, L. (Ed.). (2008). *Handbook of International Research in Mathematics Education*. New York, NY: Routledge.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1-17.

Hamilton, E., Lesh, R., Lester, F., & Yoon, C. (2007). The use of reflection tools to build personal models of problem-solving. In R. Lesh, E. Hamilton, and J. Kaput (Eds.), *Foundations for the Future in Mathematics Education.* (347-365). Mahwah, NJ: Lawrence Erlbaum Associates.

Hjalmarson, M. & Lesh, R. (2007) *Design research: Engineering, systems, products, and processes for innovation*. In L. English (Ed.), *Handbook of International Research in Mathematics Education*. New York, NY: Routledge.

Kahneman, D., Slovic, P., & Tversky, A. (Eds). (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: the Psychology of Intuitive Judgment*. New York, NY: Cambridge University Press.

Klopfer, E., Yoon, S., & Perry, J. (2005). Using palm technology in participatory simulations of complex systems: A new take on ubiquitous and accessible mobile computing. *Journal of Science Education and Technology*, *14*(3), 285-297.

Konold, C., & Lehrer, R. (2008). Technology and mathematics education. In L. English (Ed.), *Handbook of International Research in Mathematics Education* (49-71). New York, NY: Routledge.

Lesh. R. (2003a) Models & Modeling in Mathematics Education. Monograph for *International Journal for Mathematical Thinking & Learning.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesh, R. (Ed.) (2003b). *Models and modeling perspectives* [Special issue]. *Mathematical Thinking and Learning*, *5*(2&3).

Lesh, R., Cramer, K., Doerr, H., Post, T., & Zawojewski, J. (2003). Model Development Sequences. In R. Lesh & H. Doerr (Eds.), *Beyond Constructivism: A Models & Modeling Perspective on Mathematics Teaching, Learning, and Problems Solving* (35-58). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesh, R. & Doerr, H. (Eds.). (2003). *Beyond Constructivism: A Models & Modeling Perspective on Mathematics Teaching, Learning, and Problems Solving*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T. (2000). Principles for developing thought-revealing activities for students and teachers. *Handbook of research design in mathematics and science education*, 591-645. Mahwah, NJ: Lawrence Erlbaum Associates.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics.* Washington, DC: Authors.

Slovic, P. (Ed.). (2000). *The Perception of Risk*. London: Earthscan.

Slovic, P. (Ed.). (2010). *The Feeling of Risk: New Perspectives on Risk Perception.* New York, NY: Earthscan.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: the Psychology of Intuitive Judgment* (397-420). New York, NY: Cambridge University Press.

Wilensky, U. (1999). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. http://ccl.northwestern.edu/netlogo/.

Wilensky, U. & Stroup, W. (1999a). *HubNet.* http://ccl.northwestern.edu/netlogo/hubnet.html. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Wilensky, U., & Stroup, W. (1999b). Learning through participatory simulations: Network-based design for systems learning in classrooms. In C. M. Hoadley & J. Roschelle (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 667-676). Palo Alto, CA: Stanford University.

Risk—A Fundamental Condition of Doing Mathematics

Wolff-Michael Roth

University of Victoria, Canada

Jean-François Maheux

Université du Québec à Montréal, Canada

Abstract: The theme of this special issue is risk. But risk is not a common topic of investigation in mathematics education, lest it be an occasional interest in "at risk" students, generally defined as those who likely will fail at school. In this study, we are not interested in this rather limited use of the risk concept. Instead, we show that risk not only is a condition of human life generally, but also a necessity for teaching and learning mathematics. To show this, we develop the concept of risk with materials from a second-grade mathematics unit on geometry. Implications are drawn concerning the particular forms of ethics that take into account the risk incurred by teachers and students when doing mathematics in the classroom.

Keywords: curriculum and lesson planning, learning outcomes, instruction, learning, teaching, epistemology, ethics.

To Live is to Risk

> [T]o act or even to live already is to accept the risk of infamy with the change at glory. (Merleau-Ponty, 1960, p. 116)
>
> The act includes a sacrifice and a risk. . . . This is the risk: the primary word [*Thou*] can only be spoken with the whole being. (Buber, 1970, p. 10)
>
> To be together, or to be in common, therefore, is the proper mode of being of existence as such, which is the mode where being as such is put into play, where being as such *is risked or exposed*. (Nancy, 1993, p. 155)

To live means to risk. To live means to ex*per*ience; and to ex*per*ience, as the etymology of the syllable *per* indicates, means to face *per*il (Roth, 2013).[1] To act, to speak, and thereby to experience inherently means to risk because, when individuality "commits itself to the objective condition in putting itself into a deed, [it] does of course risk being altered and perverted" (Hegel, 1806, p. 258). This is especially so because what becomes of acting and speaking, inherently, is unknown and, therefore, inherently is associated with peril. Whereas it is true that "the individual, who has not dared his life, can be recognized as *person*," this person "has not attained the truth of this recognition as an independent self-consciousness" (p. 119). More recently, the very notion of agency also has been characterized in its essential character of risk: "life . . . [is] answerable, risk-fraught, and open becoming through performed actions" (Bakhtin, 1993, p. 9). Only technical rationality, the quintessence of metaphysics, cannot admit the equivalence of life and risk: both technological rationality and metaphysics presuppose the identity

[1] This aspect of human experience has found its entry into commonsense sayings as "Wer nicht wagt der nicht gewinnt [Who does not risk does not gain]" (ger.), "No guts, no glory" or "Who dares wins," "La chance sourit aux audacieux [Chance smiles at the audacious" (fr), "Chi non risica, non rosica [Who does not risk does not nibble]" (it), or "Quem não arrisca, não petisca [Who does not risk does not nibble]" (pt).

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 203-225

of knowledge and Being (*Sein*) and therefore the predictability of life (Heidegger, 2006). In mathematics education, a vast majority of curricula, textbooks, or lesson plans still are employed as technologies, as means-to-ends (Maheux, 2011): metaphysical in their assumed transcending and ordering of the messiness of life.

Metaphysics is the human entanglement in the technical and in the will (Heidegger, 2009). In metaphysics—technique and cause–effect (will) reasoning—there exists a tendency or desire to eliminate risk, and perhaps the hope that risk can in fact be eliminated. This tendency, desire, and hope is embodied by technical thinking with its striving toward "functionalization, perfection, automatization, bureaucratization, information" (p. 60). The desire to technologize life is to submit it to control: "Technopoly is a state of culture. It is also a state of mind. It consists in the deification of technology" (Postman, 1992, p. 71) and, therefore, is the epitome of has been called *the will to power* (Nietzsche, 1954a, 1954b, 1954c). Heidegger refers to it as *onto-theo-logical*, because the same desires underlie the beliefs in horoscope and pre-determination of life on the part of some deity, and the political aspirations of totalitarian states. But, as theorists increasingly recognize, technology and the modernization representing will to power are in fact sources of risk, including sources of risk to the self (Beck, 1992; Giddens, 1991).

Less ominous, the same kind of thinking exists in the practice of curriculum design, where the *What*s and the *How*s of learning are specified in advance of the mathematics curriculum as an open event of yet to be determined nature (Roth, 2013a), as if it were possible to plan mathematical learning administratively (Neyland, 2001). Because schooling as institution has the (societal) task to plan high school and college completion and the production of the qualifications for subsequent job opportunities, "learning has to be seen as something that *can be directly administratively planned*" (Holzkamp, 2013, p. 118). That is, in the institutional practice of curriculum specifications and the specific lesson plans—which determine in advance what students should know after a year in the mathematics class or after doing a task—we find the same kind of technical rationality that characterizes other parts of life in society. "Learning technologies" constitute but another means in the ill-fated attempt "'[t]o make learning more efficient and more interesting'" (Postman, 1992, p. 171).

On the other hand, we all have to recognize that doing mathematics is an expression of life. It is something we, human beings, do as part of our being-alive. In that sense, to live mathematics is to take risks, and to risk means to expose oneself to hazard or danger. When teachers and students engage in doing mathematics (together), they thus inherently expose themselves to all sorts of risks:

> [1]risk, noun
> 1: possibility of loss or injury: peril
> 2: someone or something that creates or suggests a hazard
> 3a : the chance of loss or the periods to the subject matter of a contract
>   b : a person or thing that is a specified hazard to an insurer
>   c : an insurance hazard from a specified cause or source
> 4: the chance that an investment . . . will loose value
>   – at risk
>   : in a state or condition marked by a high level of risk or susceptibility
> [2]risk, verb
> 1: to expose to hazard or danger
> 2: to incur the risk or danger of
> (from Merriam Webster Online dictionary, 2014)

Translations
Risiko (ger., neut), risque (fr., m), rizq (arab.), ρίσκο (risko, gr., neut), риск (risk, rus., m), rischio (Ital., m), ryzyko (pol., neut), risco (por., m), riziko (slv., m), riesgo (spa., m), risk, riziko (tur.)

The theme of this special issue is risk. Saying to risk, as the dictionary definitions show, is saying to expose oneself to hazard or danger, to engage the possibility of peril. Someone or something that constitutes a hazard is *a risk*. Interestingly, as the translations show, the term *risk* is one word that with little change exists in many, very different European languages. Yet risk is not a common topic of investigation in mathematics education, other than the occasional interest in *at-risk* students (e.g., Karsenti, 2010). *At risk*, according to the dictionary definitions quoted above, is used in (mathematics) education as being "in a state or condition marked by a high level of risk or susceptibility." In the context of education, *at-risk students* tend to be defined as those who likely will fail at school. In this study, we are not interested in this rather limited use of the risk concept. Instead, we argue that risk ought to be recognized as a condition of human life generally and of mathematics education specifically. From this specification arise particular forms of ethics that take into account risk as a condition of the living mathematics curriculum.

Risks, Plans, and the Living Curriculum

Many characteristics ascribed to humans do in fact exist among other animals: culture, tool use, division of labor, exchange (food for sex), and distribution (Roth, 2009a). However, whereas, for example, a bee humiliates many human builders with its precision in constructing a honeycomb, human beings uniquely demonstrate a capacity to plan ahead.[2] That is, a distinguishing feature of human beings has to do with how we appear to control our life conditions through forward thinking or, as Marx and Engels (1962) put it: "the worst builder excels the best bee in having built the cell in his head in advance of building it from wax" (p. 193). Curriculum planning is part of a technology of control over classroom events, and teachers and researchers are familiar with the ways in which mathematics curricula are specified by the relevant educational authority. Thus, for example, in the Canadian province of British Columbia, there are instructional resource packages specifying what students are expected to do (such as describing, comparing, and constructing two-dimensional shapes including rectangles and squares) along with prescribed learning outcomes and suggested achievement indicators (Figure 1). Inasmuch, curriculum planning is also controlling what counts as legitimized knowledge to be acquired (Postman, 1992), indicating what (all) students are expected to achieve (in this case: the identification of common attributes of typical two-dimensional shapes, such as rectangles, from sets of the same type and create a pictorial representation of a given shape).

Inherent in such curriculum specification—and in the associated lesson plans of teachers—is the assumption that learning can be preplanned. As a consequence, the real subject of the process begins to change: "the subject of learning would not be the pupil, but the teacher, who has to plan and perform the lessons in such a way that her/his teaching efforts are directly measurable and controllable by the effects they produce in the pupils" (Holzkamp, 2013, p. 118). Even more so when teachers are expected to treat all students the same, exposing everyone to the same curriculum, outcome differences can be and are explained by means of students' different abilities. This way of thinking about the curriculum does not account, however, for the abyss that exists between *any* plan and the situated actions that it is designed

---

[2] There is some evidence, though, that chimpanzees have such capacity, as seen in the fact that they may carry tools for 50 km from one location to another to crack nuts (Roth, 2009a).

to control (Suchman, 2007). Even the most experienced scientists think they act or have acted according to a plan only to find out later that what they have done was not what they thought to have done (Roth, 2009b). That is, there is an *eo ipso* difference between any lived curriculum and its prefixed form (Pinar & Reynolds, 1992), which evidently implies the persistent present of risk that the intended form is not realized in the living curriculum (Roth, 2013c).

SHAPE AND SPACE *3-D Objects and 2-D Shapes*	
Prescribed learning Outcomes	**Suggested Achievement Indicators**
It is expected that students will:	*The following set of indicators may be used to assess student achievement for each corresponding prescribed learning outcome.* *Students who have fully met the prescribed learning outcome are able to:*
C8 describe, compare, and construct 2-D shapes, including – triangles – squares – rectangles – circles [C, CN, R, V] [C] Communication [CN] Connections [R] Reasoning [V] Visualization	❑ sort a given set of 2-D shapes and explain the sorting rule ❑ identify common attributes of triangles, squares, rectangles, and circles from given sets of the same type of 2-D shapes ❑ identify given 2-D shapes with different dimensions ❑ identify given 2-D shapes with different orientations ❑ create a model to represent a given 2-D shape ❑ create a pictorial representation of a given 2-D shape

Figure 1. Constructed based on the student achievement specifications matching *prescribed learning outcomes* with achievement indicators from the mathematics curriculum for second grade students in the province of British Columbia. (from MoE-BC, 2007, pp. 45 & 110)

At yet another level, it is accepted within different theoretical approaches that students "learn" based on what they already "know" not in the least because their "interpretations" and "constructions" are a function of their previous experiences. In more general terms, acting (including reflecting) expands a person's action possibilities; much of this expansion is not and cannot be anticipated precisely because what will have been learned *inherently* cannot have been the intended object, because it is unknown until revealed in the course of learning, whether the subject is a second-grade mathematics student (Roth, 2012) or one of the great "poets," including Yeats, Galileo, and Hegel (Rorty, 1989). But if what we come to know is unforeseeable, then there is risk. This is so not only for the student—who cannot predict the learning outcomes are, whether s/he is on the right track to get there, and whether what s/he is doing is in fact part of the required trajectory—but also for the teacher, who is never assured that any specific lesson is going to lead some, many, most, or all students to exhibit actions that are consistent with the expected achievement indicators. But as the living curriculum unfolds, teachers and students present one another with achievements and expectations, which provide them with opportunities to navigate the inherent gap between the plans for and outcomes of the curriculum. As a result of their transactions, teachers learn and students teach to the same extent that teachers teach and students learn: all participants expand their room to maneuver (Roth & Radford, 2010). But the risk of failing still exists, even in the most well-thought and planned curriculum. This so is to a great extent because the living curriculum constitutes an unfinished event*-in-the-making whose specific nature can be determined and grasped until *after* it has ended and therefore no longer exists as event*-in-the-making (Roth, 2013a).

A recently published book-length study of mathematical learning from a cultural-historical activity theoretic perspective provides descriptions of student engagement that clearly exhibit different responses to the risk in mathematics classrooms (Roth & Radford, 2011). On the one hand, we see Aurélie engaging with the task for a while, and then states that what she is doing makes no sense, before, a few moments later, she concludes: “I don’t understand and I will never understand.” Aurélie pounds on the table and throws herself against the backrest. For the remainder of the lesson she slouches in her chair or has her head buried in her arms on the table. In any event, it appears as if she no longer engaged other than to copy what her group mate Thérèse had done. Mario, on the other hand, continues despite what looks like extreme frustration, and despite articulating failure to understand what he—unbeknownst to him at the time—is yet to come to understand. He also engages in a relation with the teacher; and, while so doing, both exhibit frustration: Mario for not knowing, the teacher apparently for not finding a way to help. In the end, however, Mario breaks through the impasse and eventually states, “Me, I now understand” (p. 89). At this point, the risk of failing has passed, the task is completed, a goal (that of understanding) is achieved: there is mastery possible only over the old and never over the radically new (Romano, 1998).

This description shows that to get to a point of understanding, a student *has to* engage with the task without knowing whether s/he will get to such a state. There is no guarantee beforehand that Mario will get to the point—in this lesson or in any lesson that follows—where he can actually state to be understanding. That is, there is an inherent risk that all the engagement, all the (intellectual, affective) investment made will get him anywhere, especially near the intended (planned) state identified in the curriculum. Not engaging, while avoiding the risk of failing in the task, actually means never reaching its intended goal. The risk of overall failure is even greater: the risk of mutism (Romano, 1998). Here, the risk is that of an inability to speak in mathematics and mathematically. In Aurélie’s statements, this risk is co-articulated, for she says that she will never understand, which, in the context, can be heard as saying that she will never understand whatever investment she makes (definition 4 above). But by not engaging, never understanding is a certainty—unless there were to be some form of immaculate and miraculous conception. From a certain perspective, the risk of failing is low; but, unacknowledged, the risk of failure is even higher at the same time. Thus, whereas speech constitutes “a conquest over silence, which nevertheless echoes in it as its most intimate risk omnipresent in all Saying” (p. 226), the making of sense has non-sense as its necessary companion.

Risks in the School Mathematics Curriculum: A Contribution to Theory Development

This study is designed as a contribution to building an epistemology that takes into account *risk* as an irremediable character of life generally and as an irremediable character of engagement in the mathematics classroom specifically. In the pursuit of this endeavor, we provide in this section exemplifying descriptions and analyses from a geometry unit for second-grade students. The lessons were to be consistent with the official, British Columbia curriculum, here with Prescribed Learning

Outcome C8 and the associated achievement indicators (Figure 1). In this section, we first provide a description of the lesson background and then present the micro-analytic dimensions of risk arising from speaking and responding; this is followed by the articulation of a meso-analytic perspective on risk at the lesson level.

The Lesson Background

The specific lesson tasks were design by a university-based mathematics educator (pseudonym Mrs. Tran) with extensive experience as a teacher at the second-grade level and the regular classroom teacher, who also served as vice-principal of the school and as a participant on governmental curriculum committees. The second-grade students represented, from the school perspective, a broad range of cultural backgrounds and ability levels including some "learning disabled" students. (Extensive background is provided in Roth, 2011a.) For the particular segment of the curriculum unit, the teachers had designed a task in which the children first used a loop of string of a specified length for the purpose of identifying all possible rectangles that could be contained within the loop given that the string had to be run along the lines of a square grid paper. Students were then to make a trace along the string to generate a pictorial representation of the shape. In this way, the students were involved in a task that would offer them opportunities to exhibit at least two of the achievement indicators that go with the prescribed learning outcome C8 (Figure 1). The two teachers showed students how they should work together to tack the string of a chosen configuration to their paper so that it would be easier to subsequently trace it (Figure 2).

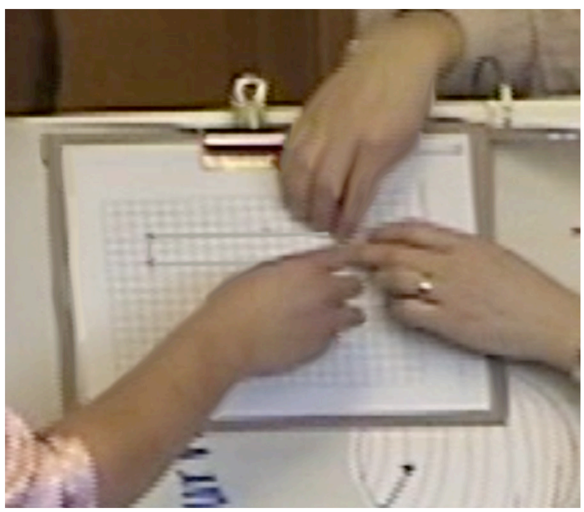


Figure 2. The two teachers—Mrs. Tran to the left, the regular classroom teacher on the right—are showing students how to lay out a string to make rectangles and fix these on a coordinate system so that they can count how many squares they contain and measure their lengths and widths

During the subsequent lesson, Mrs. Tran had different students come to the board to identify a rectangle among those available on the tray of the board that had been among the ones they had found. In the end, there were seven rectangles on the board, each containing a number that specified the number of unit squares in the rectangle (Figure 3). Using a piece of chalk, Mrs. Tran had added the lengths and widths of each rectangle in terms of the number of squares making the side. She asked students to discuss in their groups what was the same about all of the rectangles on the chalkboard or on their working sheet. After a while she brought the students back together into a whole-class setting. Even before the class had settled, she (*TT* in the transcription) was calling on Thomas, who apparently had raised his hand.[3]

Figure 3. Mrs. Tran and Thomas at the chalkboard where there are seven rectangle displayed all produced by a string with fixed length

Risk in Speaking

01 TT: ((*time on tape: 0:20:20*)) you=re listening to thomas right now ((*points towards Thomas*)) sh:: ((*places her right index finger against her mouth*)) thomas <<all, h>what is the> ↓ same about A::LL of [these ↓ RECtangles.]

[((*pointing gesture from left to right along three of the rectangles, Figure 3*))]

The common approach to the analysis of verbal transcription makes two assumptions that close analyses of speaking have not borne out: (a) the thought is preformed and merely exits via the mouth as a sort of translation mechanism and (b) the social action of a statement is given with the statement (e.g., a question is asked). Thus, a traditional analysis of turn 01 from the lesson fragment might state that Mrs. Tran is calling on the students in the class to attend to Thomas, whose turn it is right then, and that she then asks Thomas a question about an invariant property of the shapes on the board. Such an analysis, however, does not take into account the risk difference between plan and situated action: Whether we have done or said what we—however vaguely or explicitly—intended to do or say is an empirical matter (e.g., Suchman, 2007). Speaking is risking, for

> without a primary exposure to the very event of speech setting itself forth, without that opening to the new and ungraspable possibilities that grasp us in each speech by

[3] The following transcription conventions are used: all words appear in small letters; RECtangles: capitalization for emphasized parts; ((*points*)): ethnographic observations; (2.20): pause in seconds; [,]: square brackets to indicate overlap; << > >: triangular brackets to indicate voice quality, all = allegro (fast), h = high pitch, len = lento (slow); ↓: step down in pitch; sh:: colon to indicate lengthening of sound; (.) unmeasured pause; →: reference to image on the right; =: equal sign for latching sounds; punctuation at the end of speech units indicate pitch movement, "."and ";" for strongly and slightly falling pitch contour, "," and "?" for slightly and strongly rising pitch contour.

> disarming our powers over it, without that *risk, consubstantial with any genuine speech*, of sinking into silence and inability to speak, it would be impossible . . . to write a single sentence. (Romano, 1998, p. 225, emphasis added)

Traditional analysis of mathematics classroom talk thereby confuses the *Saying* with what will have been issued from it, the *Said*. While the Saying unfolds, nobody, not even Mrs. Tran, can know what she will have said once her Saying will have come to an end. This is so because "if speech presupposed thought . . . we could not understand . . . why the thinking subject himself is in a kind of ignorance of his thoughts as long as he has not formulated them for himself" (Merleau-Ponty, 1945, p. 206). At the level of the utterance, this is a moment at which risk no longer exists for the speaker: the leap is taken, the deed is done.[4] Thus, rather than investigating what Mrs. Tran will have said, that is, the finished statement in turn 01, we need to analyze the Saying as it unfolds and *before* what ultimately will have been said is available (Roth, 2013a).

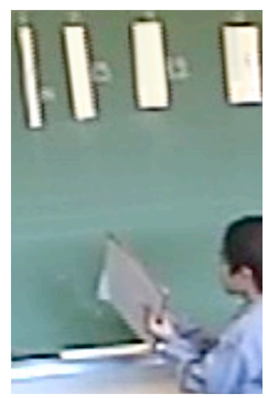
The need to analyze the Saying as it unfolds rather than the Said, which is available only when the speaker has ended and someone else speaks (acts) is quite evident from what follows because only 0.39 seconds later, Mrs. Tran continues to speak: "they all have" (turn 03). Whereas the statement in turn 01 is grammatically complete, the statement in turn 03 is grammatically incomplete. In fact, there have been suggestions that mathematics teachers produce *designedly incomplete statements* with the underlying intent that students complete these (Roth & Thom, 2009). That is, we may say that two different possibilities are offered for the nature of the next Saying. In the first, we may hear a question even thought the pitch decreases towards the end of the Saying as this is common in constative statements rather than in questions. In the second, an unfinished statement is provided, which invites to be completed especially if there were a pause.

02 (0.39)

03 they ALL ha:ve

04 [(0.46)]

[[((*Thomas walks to left of the display; gaze moves from at about second and third, then moves towards his paper* →))]



We therefore observe a change in the offering. If Mrs. Tran had an intention completely formulated in her mind and was simply externalizing it by means of her Saying, she would not have needed to make a second offering, or, alternatively what was the actual second offering would have been the only one. We see an apparent change *after* she has completed what comes to be the first offering; and it is in the completed Saying that the result of her Saying, the Said, has become available. Only then, when externalized from pure Being, are our thoughts available to us and can become an object of conscious awareness and reflection (Hegel, 1807). If Mrs. Tran had heard the pause in turn 02 as a non-response on the part of Thomas, this would then show that her intended action was inappropriate for addressing Thomas. But why would an experienced elementary teacher such as Mrs. Tran ask a question that is inappropriate for the student facing her? The answer is that Mrs. Tran, exemplifying thereby all teachers, cannot inherently know with absolute certainty whether a student does or can understand what

[4] Risk, of course, still is present, as we see below, at the conversational level.

she will have said or what s/he will have heard before she speaks, an not even once her Saying has come to an end. This uncertainty with respect to what she is doing while speaking or acting is subject to a double risk: not only the risk of non-sense referred to above, but also the risk of a sense completely other than that intended (Merleau-Ponty, 1964), as we see in the next section.

Risk in Listening

A person can reply to another person's talk only if s/he is actively listening. But listening means opening up to receive. Listeners cannot know beforehand what is coming at them. But opening up to receive includes the risk of being hurt. Recipients, in listening (or reading), do get hurt—there does not appear to be a day without another report of bullying, where a person feels hurt, insulted, or embarrassed by words someone else has voiced. Any active listening, therefore, comes with risks. Any articulation, such as a question, comes with risks that cannot be anticipated: the recipient comes to be exposed, hurt, or insulted in addition to the positive surprises that a Saying will have brought about when it has come to a close and the Said is available.

Whether the offering is taken as such remains uncertain. We do know this from the everyday interactions with others were, after intending to say or ask something, the intended addressee replies saying something like "Why are you insulting me?" or "F… off!" Now the initial speaker has to deal with the fact that what had been said has been heard as an insult independent of what a speaker might have intended. The social situation is such that the (unintended but nevertheless felt) insult has to be dealt with. What comes is heard in response to the naming of something as an insult. If Mrs. Tran had been teaching in an urban school, the student reply to turn 01 might have been "Why don't you screw someone else?" That is, *because someone is listening,* actively having the words ring in his/her ears while these are produced in her mouth (Roth, 2014), Mrs. Tran cannot ever be certain about the effect that these will have, and she cannot ever be certain about what she will have said as long as her Saying continues. The idea that thought cannot know what it is until it has externalized itself is not new but has been articulated in dialectical philosophy (Hegel, 1807) in terms of a thought (in-itself) that can grasp itself only when external to itself (where thought is for itself). More recently, dialectical psychologists continued this line of theory by suggesting that "thought is not expressed in word, but it completes itself through it" (Vygotskij, 1934, p. 269). Thought is complete only in the word.

Following Thomas' orientation towards the left end of the row of rectangles, Mrs. Tran can be heard saying, "Oh, we are talking about all of them" while gesturing along the row. It is as if she were replying to a statement about one or a few of these rectangles. In fact, one transcriber had written "these two ones" in turn 04, as if hearing something to which Mrs. Tran replied in turn 05.

04 [(0.46)]
[[((*Thomas walks to left of the display; gaze moves from at about second and third, then moves towards his paper* →))]

05 TT: oh were talking about ↓ A::LL of them ((*motions to the cards with her right hand*)) they A::LL have what.

06 (1.33)

For one, we see here how listening is not only on the side of the addressee, but also is an integral part of uttering, for the speaker not only listen to herself, but is in constant coordination with the evolving nature of the environment that her active talk is producing. The payoff associated with immediate feedback does not, however, come without risk. Conversations can quickly run off topic, apparently stall, or become paradoxical as in the famous "Topaze effect" described by Guy Brousseau, in which a student's dubious facial expression leads the teacher to answer himself the question he seems to be asking. In such a case, it is not only one "planned talk" (if such a thing ever exists) that can be compromised, but the whole teaching and learning project, especially when the teacher wishes to work from students' observations instead of monopolizing the mathematical/conceptual communicational space (Maheux & Roth, 2014). Listening to another also comes with this sort of conceptual risk, because the other's Saying is an offering that, by definition, *presents* something toward which the listener has to orient himself (should it only be by ignoring it) because it is a social fact however one stands with respect to it (Star, 1991). When Mrs. Tran says, "what is the same about all of these rectangles" and "we're talking about all of them," she opens up a space for Thomas to articulate an observation, but also offers a framing of what is to be talked about. Attending to what can be heard implies, for Thomas, the risk of having to answer to something which demotes or even contradicts his previous observations—or situates him as "not listening." There could be a case in which he was, indeed, just about to comment on two of the figures, and is prevented doing it after Mrs. Tran's emphasis on "all" the rectangles. In addition, researchers studying communication in mathematics increasingly insist on the artificiality of distinguishing between communicating with oneself and with another (e.g. Sfard, 2008). The risks of listening are fully present from the moment that something-to-be-listened-to is made available, including of course the non-verbal.

Up to now, Thomas has not yet talked, though he has walked closer to the left end of the array of rectangles (turn 04). But rather than having a turn at talk, it is Mrs. Tran who is speaking again. She makes a statement concerning the topic of the talk: "We are talking about all of [the rectangles]." She continues with a designedly incomplete statement, "They all have . . ." and inherently specifies that it is some thing or property that has to come next here marked by the earlier articulated interrogative pronoun *what*. Once Mrs. Tran has finished speaking, a pause unfolds. When it will have ended, it has been longer than the normal 0.8 seconds that teachers tend to provide on average for students to respond (Tobin, 1987), and it has been longer than the standard maximum silence in conversations of about 1.0 second (Jefferson, 1989). From the present perspective, the risks increase as the pause gets longer.

05 TT:oh were talking about ↓ A::LL of them ((motions along the rectangles with her right hand)) they A::LL have what.

06 (1.33)

07 T: they <<len>A::::LL:: HA:::v:::e;>

08 (8.30)

09 T: ((*shrugs shoulders, turns to Mrs. Tran*)) i dont know.

On the one hand, Mrs. Tran—as teachers in a study of classroom questioning revealed (Roth, 1998)—might risk putting a student on the spot, which for many gender-sensitive teachers becomes a particular issue when the turn involves a girl. Thus, for example, in that study a teacher had general guidelines for their questioning such as "Always wait 5 seconds before accepting an answer" (p. 370), but "felt uncomfortable waiting in the case of Renata and Carla, for she could 'see their discomfort, sort of squirming'" (p. 370). That teacher did not wait as long as her rule for action suggested but turned to other students rather than letting a girl "squirm." There is a risk also for the teacher, in that waiting for a student to respond fully conflicts with a rule such as *Keep a class going* (Roth, 1998). Mrs. Tran does indeed wait; and, after what will have been a short longish wait, Thomas begins to speak. In fact, we might consider him to have already begun *responding* while he attended to Mrs. Tran's speaking, and the pause is part of the production of the reply (Figure 3). Thomas draws out the words "they all have" (turn 07), which will have lasted 2.18 seconds when he ended.

Risk of Responding

If listening can be conceptualized as an early part of the response, especially when attending is made visible to the speaker, it is also possible to think more specifically about the risk involved in the responding dimension of the act (including that of attending). In the end, 45 seconds will have been taken from the first call on Thomas to Mrs. Tran's request for him to sit down and to put his hand back up when he will have thought about something that all the rectangles have. That is, Mrs. Tran will have allowed that all of that time be spent on providing for Thomas to produce a reply to question intended to state an invariant property of the rectangles only to find out that Thomas does not know—and, perhaps, given that she said "think about something that they all have," he might not have done the required thinking at all.

When he stops talking, another speaking pause begins to unfold; and, when it ends, it will have been an extremely long pause of 8.30 seconds. In the words of the teachers in another research project, Mrs. Tran really has done a lot (everything?) to let Thomas respond fully at the risk of not keeping the class going. One might say that everyone else had to wait. But if Mrs. Tran had begun a speaking turn earlier, she also would have risked not giving Thomas the opportunity to respond fully and, therefore, not providing him with an opportunity of articulating his thinking and, in this, not affording him an opportunity of learning. But in waiting for such a long time, other risks increase: (a) the possibility that Thomas does not get to reply or reply in a manner inappropriate with respect to the curriculum and (b) the possibility that Thomas will say nothing at all or admit to not knowing. It is this latter that we will have been observed once Thomas has spoken again. "I don't know," he says (turn 09).

By raising his hand, Thomas has taken risks, including not being able to provide the reply that the teacher is intending and having to admit his ignorance. Thomas is in a situation that some individuals experience as having been put on the spot, to which he contributed through his hand raising, setting up the risk that he actually would be chosen. In fact, Mrs. Tran's statement invites him to sit down, an invitation that Thomas accepts in and through his actions of returning to his seat and sitting down. More importantly, perhaps, her statement "you think of something that they all have and put your hand back up" can be heard as saying that he had not done any thinking about the question that was to be answered: "What do they all have?" or "What is the same about all of these?" By asking for a turn, Thomas has taken the risk of failing. We have no access to the emotional reflection of the situation on the part of Thomas. We do know that Mrs. Tran tended to utter something in the form of a question, wait as student were raising their hands, and then called on one of those students who had raised their hands. In the present, she had invited Thomas and, once the class was called to order, Thomas had arrived in the front of the classroom and next to the chalkboard (Figure 1). Mrs. Tran articulated a statement, which, as we see from the actions of Thomas, will have been treated as a question.

Such analysis clearly shows how responding is an expression of the *respons*ibility (response-ability) for what had been said, realizing it in a way or another, as a question or an insult and so on. Bakhtin (1986) insists that *any* utterance is a response, a link in the chain of communication created by all preceding utterances: "Each utterance refutes affirms, supplements, and relies upon the others, presupposes them to be known, and somehow takes them into account" (p. 91). Uttering or responding is thus taking the risk of contributing to continuation (or the momentary interruption) of a series of exchanges, a risk that cannot be avoided from the instant that a student or a teacher recognized him/herself as an addressee. When Mrs. Tran says "they all have what" and Thomas responds, "They all have . . . I don't know," she makes available to him a verbal utterance to which he finds himself in the obligation to answer even if he does not know what to say, as his answer shows (the hearable hesitation and the concluding "I don't know"). We are not saying that that the teacher riskily puts her student on the hot seat, but that there is an inherent risk that comes with responding, and that any utterance has the potentiality to trigger a response if someone (including oneself) recognized him or herself as its (potential) addressee. In as much, Thomas's final words in the previous fragment, his "I don't know," constitutes, in turn, a statement that offers itself to Mrs. Tran (and other people in the room) to be responded to, a statement that places everybody in presence at risk.

In this particular instance, the risk of Thomas's response is perhaps even higher since the topic of the last few verbal exchanges involve "knowing" about 2D shapes at the end of a lesson where students could be expected to notice regularities, thereby demonstrating the success in what the teacher had planned. The whole lesson is somehow at stake through his admittance of not knowing, and Mrs. Tran's response to Thomas's affirmation can lead to very different outcome, whether she manages to overcome this impression of knowing or not-knowing. We all know instances where children leave a class saying things like "it was so confusing today . . . we didn't learn anything . . . why did we even do that?" We also know instances, where a teacher makes just the "perfect, magical" comment that allows all the pieces of a puzzle fall into place, and makes students feel like they really got to learn something that day.

Risk of Conversing

As we can see, teacher-student exchanges present inherent risks in a number of ways. First, what might be intended as a question may in fact expose a student, if he has to admit not knowing and, thereby, be exposed as not knowing (Maheux & Roth, 2014). There are reports that teachers are especially sensitive when their questions address girls (Roth, 1998). In a fourth-/fifth-grade science class, the two teachers directed four times more questions to boys than to girls in whole class discussions because they did not want to expose the latter. For example, one of the teachers told a boy to "'shut up' as part of her attempt to provide a safe environment for girls to engage in risky responses" (p. 368). Second, Mrs. Tran cannot not know whether asking Thomas to come to the front of the classroom is taking the lesson any further. What is it that their transactional exchange will have produced once they are done and Thomas is asked to return to his seat? What if the two were to continue and thereby produce a lengthy exchange only to be confronted—after everything has been said and done—with the fact that nothing articulated has moved the lesson any further in terms of the curriculum plan?

Conversing is realized the articulation of utterances whose outcome (including their meaning, and contribution to mathematics teaching and learning) are also produced in the flow of uttering/responding. As such, conversations also include the notion of *time*, where risk is situated both in the here and now of taking a leap "after all that happened," and in the projected open space of "all that can follow." In our episode, for example, the making and taking of time is apparent in the particulars of turn 07. That is, in the drawing out of the syllables, time is made, which makes it possible for other things to be prepared. In that same action, Thomas is taking (his) time to produce the two parts of the reply both in the pause and the production of the words. In this double belonging of the time to Mrs. Tran and Thomas, we are provided with an example of the idea of *social* action that cannot be reduced to individual action: any pause, as much as any single turn (see below), is to be theorized as a joint rather than an individual action. This leads us to a revised model of talk in which any part of a conversation has sociological and psychological, synchronic and diachronic dimensions (Figure 4). It is because each word belongs to a minimum of two people (Vygotskij, 1934) that signification itself is at stake, where sense always exists over against non-sense into which it may collapse. There are no definitive a priori rules or laws that specify the sense of an unfolding Saying (Davidson, 1986); any sense *advenes* in the event of speech and therefore cannot be reduced to it (Romano, 1998).

The analysis of classroom talk focuses on speakers and tends to forget that at the very instant that a speaker speaks a listener listens. That is, every word exists twice, simultaneously: for the speaker and recipient. Our analysis has to take this into account (Roth, 2014). This simultaneous (synchronic) presence of the word is captured in a revised model of the exchange, where it constitutes the sociological dimension of speech (Figure 4). On the part of the recipient, *actively attending* to the other constitutes the first part of the response; the second part is comprised by the actual *reply* (Figure 4). In the unfolding of the response, there is actually a transformation where, from the respondent's perspective, the Saying of the other is transformed into a reply. This transformation constitutes the psychological dimension (Vygotskij, 2005); this dimension also is diachronic. It is not so much that the response occurs in time, but that responding is at the very origin of time: it is making and taking time. This aspect of the exchange is particularly visible in the speaking pause, which makes space for either participant to take the next speaking turn, and for the respective other an opportunity to take the listener's turn. Who knows whether this time is not going to be wasted? Who knows whether the current part of an exchange will not have been putting one or the other participant in a situation where s/he lost or risked loosing face?

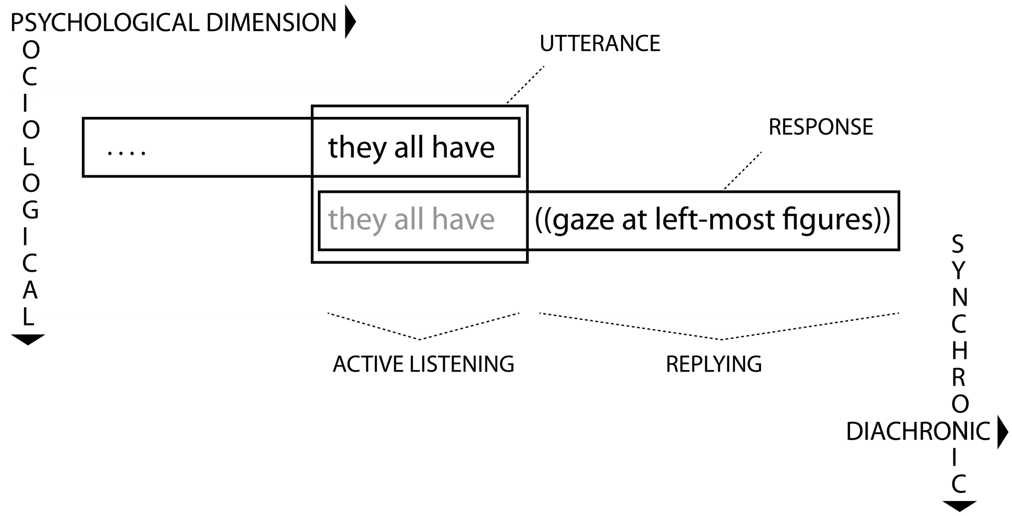


Figure 4. Every utterance is utterance only because it is both said and received; every response consists of active listening and replying.

Many philosophers agree that each act, including its negation the non-act, constitutes a risk (e.g., Bakhtin, 1993). In the preceding subsections, we exemplify risk as an essential dimension of each individual mathematics-oriented act and conceptualize the act in such a way it they cannot be reduced to the sum total of independent, individual locutions. Each action is the result of a dialogical interplay of turns in conversation, part of a dialogue (literally: "two-voices") demanding that in each turn, both parties are involved: acting (speaking, raising hands) and attending (active listening, watching) crossing over into one another. Such acts are integral and constitutive parts of classroom episodes that have distinguishable beginnings and endings. For example, a student's turn is one such episode (Roth & Thom, 2009). In the lesson fragment drawn on here, it is Thomas's turn, which consists of a sequence of turns at talk. The outcome of the enchainment of *joint actions* that constitute the overall turn multiplies the risks. It is precisely the dialogical nature of the relation, here between teacher and student, that makes the outcome uncertain so that even novelist cannot know where the narratives are taking them (Bakhtin, 1984a). Thus, as we write above, when Mrs. Tran calls upon Thomas she cannot know whether what will have been said has taken the lesson further towards achieving what the lesson plan has laid out.

Once and when Mrs. Tran is heard to be asking a question about what *all* the rectangles have in common, that is, once the content and the purpose of the statement becomes definitive, students' past and present perceptions are addressed. What is it that all of the rectangles on the chalkboard have in common? It is based on what students see—their perceptions—that students are expected to and will respond. This introduces a further dimension of our phenomenon because perception itself means risk: for it "would not know to win on one side without losing on the other and without exposing itself to the risk of time" (Merleau-Ponty, 1945, p. 396). Perception, then, rather than taking in the real and objective, means gaining and losing simultaneously, being exposed to the risk coextensive with life.

In the above-mentioned study in a fourth-/fifth-grade science course, even teachers with considerable experience at the grade level and substantive content related knowledgeability do not inherently get the timing of questions right or know when to stop a line of questioning (Roth, 1998). Teachers who are familiar with the students in their care have a better "sense for drawing the line between productive questioning and discouraging a child because their previous efforts [at soliciting an answer] appeared insufficient" (p. 364). Despite her considerable experience as a classroom teacher and despite her substantive knowledge of mathematics and mathematical learning as a professor of mathematics education, Mrs. Tran's questioning risks being inappropriate for the particular child in front of her (here Thomas). This inappropriateness may lead to exposing the child and to negatively affect his/her current emotional state specifically and the overall development in general.

Risk in the Mathematics Lesson

The dictionary definition of *risk* provided above includes the someone or the something that creates or constitutes a hazard. From the perspective of this definition, the British Columbia curriculum may be considered a risk because it is a setup for failure. Students may fail to exhibit the prescribed learning outcomes and teachers may fail to exhibit what teacher educators and researchers of teaching call *pedagogical, subject matter pedagogical,* and *general pedagogical knowledge*.[5] In the preceding subsection, we present aspects of risk at the (micro-) level of individual and collective action, which really is the risk arising from living and relating as such. In this subsection, we focus on the teacher and the risks incurred in the very attempt to realize a previously articulated lesson plan. This is a meso-level analysis when we consider the particular goal of Prescribed Learning Outcome C8 in the context of the second-grade mathematics curriculum as a whole, and the effects of school mathematics on students in its entirety (e.g., loss of interest, completion of applied vs. academic level mathematics courses, opting out of mathematics-related careers). Here, at the meso-level, the curriculum was to get students to identify that the rectangles on the board all have the *same* perimeter, while *differing* in terms of the shape—long and skinny, almost square—and in terms of the number of unit squares they contain (i.e., area). Each time Mrs. Tran calls on a student to make a contribution generally and to respond to an apparent query—*What is the same about these squares?*—specifically, she incurs the risk of failing to receive the desired reply to the specific question at hand and to move closer to the intended end of the lesson at which point students should be doing what the curriculum specifies: describing similarities; describing and comparing (a set of) rectangles (Figure 1).

Following Thomas, Mrs. Tran calls on Lucas, again running the risk that what she is asking for, according to her lesson plan, is not going to be provided in and by the reply. Like Thomas, Lucas is

[5] It may not come as a surprise to find many teachers apprehensive of observation and averse to participate in research.

making and taking his time, saying that the shapes are rectangles and, thereby, stating what the instructions had started out with: make different *rectangles* using the string.

10 TT:okay you sit down ((*motions to the desk*)) you think of something that they ALL have you put your hand back up. ((*Thomas walks to his seat*)) who can think of something that they ALL HA::VE ((*motions to the cards with her right hand*)) or that they ALL ARE. (1.53) lucas

11 L: u:::::h (0.81) they are all rectangles,

12 (0.36)

13 TT:they are all rectangles yes, and why are they all rectangles ((*motions to the cards on the board*)) lucas

14 L: u::::::h because (0.84)) they (1.69) they kinda width is lo:nger and kinda skinny

15 TT:and so if theyre not all the same size around ((holds her hands a distance apart side by side and one on top of the other)) or the length around them we call them oblongs ((points to a rectangle)) right lucas, so theyre all oblongs ((*points to the rectangles*)) theyre all rectangles but theyre all oblongs because none of them are squares (.) what is else is the same about all of them ((*motions to Melissa*)) melissa

Although she is talking a lot, the exchange with Lucas has not taken the lesson any further. But Mrs. Tran is not abandoning. She is giving it another try, this time asking Alycia, and thereby taking additional risks that the sought-after reply will not be forthcoming and that the time invested will have no return.

16 M: um that theyre alls that theyre all rectangles

17 TT:yes they are all rectangles yes theyre ALL oblongs and theyre ALL rectangles ((*motions to A*lycia)) alycia

18 A: um they all have four edges and verticies

19 TT:they all have four edges or four sides and they ((*places her right pointer finger along the outside edge of a card*)) all have four verticies or four corners and heres something else (.) they ALL use ((*holds her hands a distance apart*)) the same did we use different loops of string, (.) ((*moves her hands closer then farther apart*)) when we made this, (.) did we did i give you a different did you use a different loop of string ((*moves her right hand in a circle*)) for each rectangle that you made, ((*motions to the board*))

Again, while acknowledging the reply and the veracity of the contents with the actual state of affairs, her continuation shows that she is still after a specific something. She makes a statement that can be heard as a question, “They all use the same . . . did we use different loops of string when we made this?” and she continues, “Did we, did I give you a different, did you use a different loop of string for each rectangle you made?

Here we also have another clear indication for the fact that thought is not finished and then made available to the audience as if in a memory dump. Instead, there are what we can hear as repeated beginnings: “they all use the same,” “did we use different loops of string when we made this,” “did we,” “did I give you a different,” and “did you use a different loop of string for each rectangle that you made?” We observe here five beginnings, and only the last one completed what we can hear as a question concerning the length of the string and, as a person savvy in mathematics curriculum can hear, implicitly about the perimeter of all the rectangles. Mrs. Tran then receives a first answer, which will have been the beginning of the end of this final push towards the desired articulation. Many students reply in unison with a "no," which sets Mrs. Tran up to take another step in producing what will have been a designedly open query: "They were all made with a," which many children replied to by saying, again in unison, "same" (turn 16). Rather than engaging in the risk of failing to receive the sought-for answer once again, she states what it seems she might have been looking for all along: very *different* rectangles were made using the *same* string, different in terms of their shape (more or less skinny).

20 TT:did you use a different loop of string ((*moves her right hand in a circle*)) for each rectangle that you made, ((*motions to the board*))

21 Ss: no:: ((*some children*))

22 TT:you you all they were all ((*points to the chalkboard with her right hand*)) made with a

23 Ss: same ((*some children*))

24 TT:same piece of string they were all made with the same piece of string ((*places her right hand to her mouth*)) yE:t we made VERY DIFferent rectangles ((*moves her right hand over the rectangles*)) (.) from the sa:me piece of string these some of the rectangles that weve made ((*runs her finger over the first narrow rectangle*)) were long and skinny and then other ones ((*runs her hands under the other rectangles*)) werent so long and skinny and they were more like squares but they werent perfect squares ((*motions to Ch*ris)) chris ((*time on tape: 00:24:39*))

Mrs. Tran has now almost entirely formulated what appears to have been the observation that she has expected all along: the rectangles have the same perimeter since they where all made using strings of the same length. From turn to turn, narrowing the conversation by means of what look like open-ended questions to a pattern in which students provide straightforward single word answers ("no", "same") was part of the risk of engaging in the lesson in that particular way. But there is also richness inherent to all instances of elocution by means of which there is always more in a saying than what is apparently said: We always say more than we can say in so many words (Garfinkel, 1967). There is always a possibility for reinterpretation and elaboration what was really meant, which Mrs. Tran's last statement exemplifies. From the students' comments about the being rectangles, she articulates an observation about none of them being squares, which was not part of the lesson's plan. And in so doing, the teacher does not merely put an end to teaching and learning, but keeps offering something to respond to, even though it is not in the planned form or with the panned content.

The lesson goes on for an additional 8 minutes without some form of statement of potential mathematical interest, such as different areas of rectangles with the same perimeter or the lengths and widths add up to 30, or one length and one width add up to 15. It might have been interesting to find out what the children would have said if asked what the take-home message of the lesson was, what the lesson had shown. This could have made it apparent to the teachers whether they had met their curriculum goal or whether their investment was a loss (write off). Without somehow getting the kids to express themselves, the teachers cannot know how close they have come to the intended goal of this day's journey. Mrs. Tran and the regular teacher of the class had planned the different tasks to achieve—in a more innovative than normal way—the second-grade curriculum specified in the official curriculum of the province. Despite all their planning, neither one could know in advance whether what they were doing on the basis of the specified plan for the lesson will succeed in the way anticipated. That is, there is an inherent risk that individual parts of the lesson or the lesson as a whole fails in that some or all students will not exhibit what they are intended to exhibit (e.g., consistent with the suggested achievement indicators).

Towards a Risk-Oriented Epistemology

The purpose of this study is to contribute to the building of epistemology generally and to a theory of mathematics education specifically that take into account *risk* as an essential feature of life. Underlying educational uses of institutional curricula and teacher lesson plans with specific learning outcomes are part of a technological rationality (Postman, 1992) of Greco-Roman philosophical origin and tradition (Heidegger, 1982) according to which mathematics students are to learn what has been set out beforehand. The purpose of plans is to eliminate, decrease, or limit risks. In the meeting with the other, in saying the word *Thou*, human beings, as the second and third introductory quotations of this text show, submit themselves to risk and exposure. Human experiences turn out to be in excess of all technological rationality, which has led to the emergence of the concept of risk, a term that has its etymological origins in the post-classical Latin *resicum, risicum* in the sense of hazard and danger, itself with possible origins in the Arabic *rizq*, *rēsq* fortune destiny, luck, and chance; these origins are common to many modern languages (see above).

In this study, we suggest that teaching and learning mathematics is characterized by risk as an essential (i.e., irremediable) feature that undermines any technical rationality of curriculum planning. What students learn and what teachers teach is unavoidably uncertain. It remains an empirical matter that can be established with certainty only after the fact—hindsight, as the vernacular saying goes, is 20/20. Life in general and the living mathematics curriculum specifically are not of this nature. In everyday mathematics lessons, students and teachers not only are subjects of activity but also are inherently subject to and subjected to conditions that are also of their making. In the Saying, the one exposes himself/herself to the other (Levinas, 1978) so that, thereby, mathematics teachers and students take and are subject to the risk of failing to achieve the previously stated curriculum (lesson) outcomes (i.e., to non-sense [Merleau-Ponty, 1964]), being hurt. In fact, without risk, there is no teaching and learning; and without teaching and learning, there is no risk.

We use episodes from second-grade geometry lessons to exhibit the inherent unpredictability of the course of the lesson and, therefore, the risk that teachers and students face. The episodes exemplify that teachers and students always act in the face of uncertainty. We show how risk in part derives from exposure to the other, which has implications for mutual and collective responsibility. Teaching, therefore, may also require a *precautionary principle*, much like other phenomena in a risk society. The notion of risk leads us to a rethinking of the subjects of/in mathematics, no longer persons actively constructing themselves, their knowledge, and their situations but *advenants* exposed to the perils of their adventures, which they can only witness but never grasp (Romano, 1998).

One important aspect of a risk-sensitive theory of acting in mathematics concerns the subject. In traditional theories (including constructivism and enactivism), the subject of activity is theorized as an agent who creates (constructs) himself/herself in the autopoietic act of creating the object. However, the subject of experience does not know what s/he will know after going through the experience, and, therefore, also is the *patient* of experience, subject to and subjected to the environing conditions (Dewey, 1934/2008; Vygotskij, 2001). This has led to the articulation of the subject as *advenant*, as someone to whom something advenes. It means that we never are the constructive agents of our experiential selves, as other epistemologies suggest, but we are, passing from self to self, exposed to the risk undoing ourselves in the face to the absolutely other: the event that we can grasp only after it will have been completed (Roth, 2011b). The epistemology we are working towards would have to include the subject as advenant and patient—capable of being affected, that is, of passibility—in addition to its role of agent and, therefore, agency. As advenant, the subject to whom something happens, the person cannot anticipate the results of actions, his/her own and those of others. When Thomas raised his hand, he could not foresee that the events would put him into a position where he not only did not supply the apparently desired reply but also where his not-knowing was exposed in front of an audience with the possibility of embarrassment. He was put on the spot, not by Mrs. Tran, not by his actions, but by the encompassing situation as a whole that transcends any individual actor present. And similarly, Mrs. Tran could know that she would have to deal with this situation, or the subsequent chain of questions and answers out of which she apparently produced the expected observation—in a text-book case of Jourdain and Topaze effects.

Curriculum planning and the specific learning outcomes that the plans state presuppose the possibility to administratively plan learning. However, if the lesson unfolds in unpredictable and unforeseeable ways, that is, if the learning process takes all sorts of routes, we cannot be certain about what students come to understand, if they come to understand anything at all. This is exemplified in the empirical materials presented above: despite a considerable number of students offering their attempts at replying to Mrs. Tran's question, she ends up articulating it herself. In the course of the students' and teacher's mutual engagement, the signification of the talk is at stake as much as the significance of the array of rectangles visible on the chalkboard (Figure 3). Here, then, the promise of sense that comes with engagement in school mathematics tasks has the non-sense of mathematics as an inherent and necessary companion. Without non-sense, sense could not be distinguished; but the search of sense occurs at the peril of remaining in, and falling back into, non-sense.

In this study we highlight the inherent risks in teaching/learning situations. These risks are not avoided when teachers take complete control over the lesson, because this might fundamentally let teacher talk unfold closer to the way it is planned with the loss of control over what students will take away from the lesson. Acting is exposing oneself to risk—of failure, of getting injured and hurt. Speaking, too, is associated with failure and injury. Speaking, by means of which thinking *realizes* itself (i.e., becomes something of the real world), also is exposition (ex-peau-sition, as one may say or hear in French, outside-of-the-skin-position), and because of the very outside of the position of self, it, too, may get hurt. The detailed analyses exhibit at which point there are risks, such as not producing a reply or even the effect that a statement will have had once we hear the reply. In the very act of listening, too, we face risks, for listening means opening up to something that we do not know in advance, with the possibility that the what (e.g., name) or how of the statement (e.g., ironic, sarcastic) being at the origin of injury, pain, hurt, and suffering. We are subject to conditions that are unpredictable even though we contribute to their making. That is, we are both subjects of the conditions, producing them, as much as being subject to and subjected to them. We see this occurring at all levels of our analysis: individual statement (turn), the episode (Thomas's turn), and the lesson. Both agential and passive aspects are constitutive of risk: Acting is risking. Acting is risking even if we are alone, and it is especially risky when we are with the other, when we address ourselves to the other (Buber, 1970; Nancy, 1993).

Above we note that in (mathematics) education, the notion of risk is generally reserved for students who, because of their class, socio-economic background, race, or gender, experience higher risks of failing than other subgroups of the student population. This study provides empirical examples for the fact that every participant in the educational enterprise is at risk, which in fact is a condition of life (Bakhtin, 1993). It is therefore better to think risk as differential: all students and their teachers are at risk. This means that we can understand how even the most highly achieving students may still not arrive at making sense, being subject to the illusion of sense in the face of apparent non-sense. The very attribution creates another risk: students, parents, and teachers believing that sense has been made when in fact there is nothing but non-sense. It therefore comes as little surprise when even and especially the normally highest achieving students approach their teachers asking questions such as "Am I right so far?," "Is this okay?," or "Am I on track?" As research among older science students has shown, they may do so even and precisely when completely buying into a (radical, social) constructivist paradigm of learning, which means that they may come to think about certain phenomena in ways that will not get them past the next hurdle posed by a (standardized) test (Lucas & Roth, 1996). Precisely because they cannot know what it is they are expected to know after the specified curriculum will have been covered, students cannot know whether what they are doing is somehow getting them there. Only after the fact might they be in a position to state that they have, indeed, arrived where they were intended to arrive, and thus appreciate why they have done what they have done to get to this point of arrival (Roth & Radford, 2011). But in asking their questions, students can, in many cases, be aware of the risks involved in engaging in curriculum tasks: loss of value of investment, possibility of peril to themselves, and being at risk. Not only their understanding of the current mathematics topics or their understanding of mathematics more generally is at risk: their future is at stake (thus, high-stakes testing), making it intelligible that they seek to minimize the risks.

One common practice to deal with the failure of curriculum is to seek fault in students and to blame *them* when the prescribed learning outcomes have not been achieved. Lack of motivation, lack of intelligence (ability), low socio-economic status, family situations (e.g., single parent family), disinterest, and other concepts are used as resources for explaining why an individual lesson or a curriculum as a whole has not led to the planned, prescribed learning outcomes. Less frequently, though still common, teachers are blamed by students, their parents, or educational critics as the reason for low student achievement. In fear of failing to reach the prescribed learning outcomes, many teachers are calling the BC Ministry of Education asking for specific descriptions for how to teach a part of the curriculum (S. Boutonné, personal communication, undated), thereby constituting themselves as technicians following plans for which they are no longer responsible (e.g., Neyland, 2001). This is but one strategy for engaging in risky behavior while securing the possibility for attributing the failure to a source of risk other than one's actions.

This study makes apparent why we require an epistemology that accounts for risk. Our exemplifying empirical materials show that risk exists at many different units of analysis one might choose for looking at mathematics lessons. This is so because the expansion of one's grasp and control over conditions, learning, inherently means risk. Thus, those who want to know from their own experience how a conqueror or discoverer of ideas feels are in need of "*the great health*—one that is such that one cannot only have but also continuous has to acquire and has to acquire, because one constantly exposes it, has to expose it" (Nietzsche, 1954b, p. 258). Nietzsche finds similarities the sciences that predict the future with certainty, the seers who know the future, and god-inspired individuals; in his view, they all are examples of metaphysic endeavors, believe that there is an underlying plan that pre-ordinates everything that can be. The world is not this way, but in continuous renewal through death of the old (ideas) and birth of the new (Bakhtin, 1984b). On this path towards new ideas, we are at risk and we in fact have to expose ourselves to it: there is no other path to the new, which inherently lies beyond the horizon of the unknown, the unseen and therefore unforeseen. In schools, as in academic life more generally, some individuals appear to be willing to take more risks than others, who are not willing to do so. In this study, we suggest that risk is an inherent feature of life generally and of mathematics lessons specifically. We all take and are subject to risk: all of the time. In this study, we repeatedly suggest that acting means risking, and not acting might mean risking even more, to the point that failure becomes certain (the above-noted case of Aurélie). Without participation in mathematics lessons, without engagement in the tasks, and without the associated possibility that the investment of time and emotion will lead to affectively negative outcomes, subjects have no hope in expanding their room to maneuver in the field of study.

Towards an Ethics of Risk in Mathematics Education

To ex*per*ience, as we suggest, means facing *per*il: the risk of failing. The open-endedness of events and experience implies a different kind of ethics than the universal ethics arising from Kantian constructivism (Roth, 2013b). In this section, we outline the issues that an ethics sensitive to risk involves.

In his critique of technocratic mathematics education, Neyland (2001) differentiates technical accountability and ethical responsibility. On the one hand, teachers are externally contrived to behave according to curricular requirements, to follow plans, so that they will be technically cleared from their students' potential failure (or success). In terms of risk, we might say that framing teachers' work in such a way aims at reducing the risk for both teachers and students, asking them to teach and learn according to a mechanistic metaphor where success or failure depends less on individuals' work that on the system as a whole (the tuning of which belongs to the technocrats). Risk is not welcomed in outcome-based curricula: It is, at the best, accepted as inevitable. On the contrary, ethically informed teaching and learning in the spirit of Levinas (1978) hinges on an acknowledgement of risk and uncertainty over an obsessive search for control. Teachers and students are received as inherently ethically responsible for the other prior to and over any form of curricular requirements. In curricular terms, this call for a rethinking of teaching and learning plans in favor of open-ended topics for mathematical investigation. This is a turn that does not exclude students becoming experts of certain mathematical ideas and techniques, but does not organize classroom around that (Neyland, 2010).

Moment-to-moment analysis of teacher-student actions in mathematics lessons from such an ethical perspective then illustrates how, for example, bringing about mathematical concepts or ways of doing have their ethical roots in the dialogical relation of participants in conversation (Maheux & Roth, 2014; Roth, 2013b). When Mrs. Tran insists (through repetition and emphasis through prosody) upon children (and Thomas more specifically) to say something about "*all* the rectangles," she takes up the ethical challenge and offers a topic, a place for dwelling in thinking. More so, the offer is clearly marked as being *for* them, and the question to be investigated *with* her. There is a many-folded risk in doing so, which includes the risk of alienating the other, of missing the ethical possibility of being there with and for the other (prior to curricular "prescribed learning outcome" supposedly demonstrated when students "identify common attributes of triangles, squares, rectangles, and circles from given sets of the same type of 2D shapes" (Figure 1). But this is *also* an exposure, an opening, a possibility for Self and Other to meet in forms of knowing characteristic of mathematics as a culturally and historically developed praxis through which past and distant others also become present (Husserl, 1976).

Mrs. Tran and Thomas responding to one another exhibit being responsible *with* and *for* the other and the condition of being as being-with in even the most banal moment of a mathematics lesson. They exemplify the ethical risk and possibility of doing mathematics with and for an other. And they also permit us to appreciate how mathematics lessons, talk, and statements are risky objects-in-the-making, always open to bifurcations and reinterpretation and so on. To be clear, at the heart of an ethics sensitive to risk—where to experience means facing peril—what matters most is *not* the assessment of teachers actions as best, or good, or appropriate practice. What matters is *not* to decide whether or not Mrs. Tran was "right" to question and answer her students in that way, or if the risk of telling them what she might have hope to hear them say was "worth it." Instead, we conceptualize curriculum as pure mobility, and lessons or utterances as ever-changing moments of events*-in-the making, with ending and outcome yet-to-be-achieved and inherently uncertain and unpredictable (Roth, 2013a). This entails that there cannot be, ever, a final statement or appreciation of human actions, in either objective or subjective terms. Ethics cannot be injected into mathematics lesson plans or preparations, and it cannot be secured. Ethics is at the tip of every single act, always in the making, dialogical in form and content, always in the coming (Roth, 2013b). For us, it means that growing awareness to the inherent and fundamental presence of risk in teaching and learning is an essential aspect in the advent, in, of, and for mathematics education, of an ethics sensitive to risk as a necessary condition of doing mathematics.

Acknowledgments

The data for this study were collected with support from the Social Sciences and Humanities Research Council of Canada. We are grateful to all participants, our mathematics education colleague, the classroom teacher, and the children.

References

Bakhtin, M. M. (1984a). *Problems in Dostoyevsky's poetics.* Minneapolis, MN: University of Minnesota Press.

Bakhtin, M. M. (1984b). *Rabelais and his world.* Bloomington, IN: Indiana University Press.

Bakhtin, M.M. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press.

Bakhtin, M. M. (1993). *Philosophy of the act*. Austin, TX: University of Texas Press.

Beck, U. (1992). *Risk society: Towards a new modernity*. London, UK: Sage.

Buber, M. (1970). *I and Thou*. New York, NY: Touchstone.

Dewey, J. (2008). *Later works vol. 10: Art as experience* (J.-A. Boydston, Ed.). Carbondale, IL: Southern Illinois University Press. (First published in 1934)

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age.* Stanford, CA: Stanford University Press.

Heidegger, M. (2006). *Gesamtausgabe. I. Abteilung: Veröffentlichte Schriften 1910–1976. Band 11: Identität und Differenz* [Complete edition. Part 1: Published writings 1910–1976, vol. 11: Identity and difference]. Frankfurt/M, Germany: Vittorio Klostermann.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166–196). Clevedon: Multilingual Matters.

Hegel, G. W. F. (1807). *System der Wissenschaft: Erster Theil, die Phänomenologie des Geistes* [System of science: Part 1, the phenomenology of spirit]. Bamberg, Germany: Joseph Anton Goebhardt.

Heidegger, M. (1982). *Gesamtausgabe. II. Abteilung: Vorlesungen 1923–1944. Band 54: Parmenides* [Complete edition. Division 2: Lectures 1923–1944., vol. 54: Parmenides]. Frankfurt/M: Vittorio Klostermann.

Heidegger, M. (2006). *Gesamtausgabe. I. Abteilung: Veröffentlichte Schriften 1910–1976. Band 11: Identität und Differenz* [Complete edition. Division 1: Published writings 1910–1976, vol. 11: Identity and difference]. Frankfurt/M, Germany: Vittorio Klostermann.

Heidegger, M. (2009). *Gesamtausgabe. III. Abteilung: Unveröffentlichte Abhandlungen, Vorträge—Gedachtes. Band 76: Leitgedanken zur Entstehung der Metaphysik der neuzeitlichen Wissenschaft und der modernen Technik* [Complete edition. Division 3: Unpublished essays, presentations—Thoughts. Vol. 76: Leading ideas to the origin of metaphysics, present day science, and modern technology]. Frankfurt/M, Germany: Vittorio Klostermann.

Holzkamp, K. (2013). The fiction of learning as administratively plannable. In E. Schraube & U. Osterkamp (Eds.), *Psychology from the standpoint of the subject: Selected writing of Klaus Holzkamp* (pp. 115–132). Basingstoke, UK: Palgrave Macmillan.

Husserl, E. (1976). *Husserliana Band VI: Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie. Eine Einleitung in die phänomenologische Philosophie* [Husserliana vol. 6. The crisis of the European sciences and transcendental phenomenology. An introduction to phenomenological philosophy]. The Hague, The Netherlands: Martinus Nijhoff.

Karsenti, R. (2010). Nonprofessional mathematics tutoring for low-achieving students in secondary schools: A case study. *Educational Studies in Mathematics, 74,* 1–21.

Levinas, E. (1978). *Autrement qu'être ou au-delà de l'essence* [Otherwise than being or beyond essence]. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lucas, K. B., & Roth, W.-M. (1996). The nature of scientific knowledge and student learning: Two longitudinal case studies. *Research in Science Education, 26*, 103–129.

Maheux, J.-F. (2011). Epistemological issues in the educational use of technology. *EDULEARN11 Proceedings* (pp. 495–505). Online: http://library.iated.org/view/MAHEUX2011EPI

Maheux, J.-F., & Roth, W.-M. (2014). The relationality in/of teacher-student communication. *Mathematics Education Research Journal*, *26*, 503–352

Marx, K./Engels, F. (1962). *Werke Band 23: Das Kapital* [Works vol. 23: Capital]. Berlin, Germany: Karl Dietz.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception* [Phenomenology of perception]. Paris, France: Gallimard.

Merleau-Ponty, M. (1960). *Signes* [Signs]. Paris, France: Gallimard.

Merleau-Ponty, M. (1964). *Le visible et l'invisible* [The visible and the invisible]. Paris, France: Gallimard.

Merriam Webster Online Dictionary. (2014). Accessed July 31, 2014 at http://www.merriam-webster.com/

Ministry of Education, Province of British Columbia. (MoE-BC). (2007). *Mathematics K to 7: Integrated resource package 2007*. Victoria, BC: Author.

Nancy, J.-L. (1993). *The birth to presence*. Stanford, CA: Stanford University Press.

Neyland, J. (2001). *An ethical critique of technocratic mathematics education: Towards an ethical philosophy of mathematics education*. Unpublished PhD Thesis, Victoria University, Wellington, New Zealand.

Neyland, J. (2010). *Rediscovering the spirit of education after scientific management*. Rotterdam, The Netherlands: Sense Publishers.

Nietzsche, F. (1954a). *Werke in drei Bänden: Band 1* [Works in three volumes: vol 1]. München, Germany: Hanser.

Nietzsche, F. (1954b). *Werke in drei Bänden: Band 2* [Works in three volumes: vol. 2]. München, Germany: Hanser.

Nietzsche, F. (1954c). *Werke in drei Bänden: Band 3* [Works in three volumes: vol 3]. München, Germany: Hanser.

Pinar, W. F., & Reynolds, W. M. (1992). *Understanding curriculum as phenomenological and deconstructed text*. New York, NY: Teachers College Press.

Postman, N. (1992). *Technopoly: The surrender of culture to technology*. New York, NY: A. A. Knopf.

Rorty, R. (1989). *Contingency, irony, solidarity*. Cambridge, UK: Cambridge University Press.

Romano, C. (1998). *L'événement et le monde* [Event and world]. Paris, France: Presses Universitaires de France.

Roth, W.-M. (1998). Science teaching as knowledgeability: a case study of knowing and learning during coteaching. *Science Education, 82*, 357–377.

Roth, W.-M. (2009a). Cultural-historical activity theory: Toward a social psychology from first principles. *History and Philosophy of Psychology Bulletin, 21*(1), 8–22.

Roth, W.-M. (2009b). Radical uncertainty in scientific discovery work. *Science, Technology & Human Values, 34*, 313–336.

Roth, W.-M. (2011a). *Geometry as objective science in elementary classrooms: Mathematics in the flesh*. New York, NY: Routledge.

Roth, W.-M. (2011b). *Passibility: At the limits of the constructivist metaphor*. Dordrecht, The Netherlands: Springer.

Roth, W.-M. (2012). Mathematical learning: the unseen and unforeseen. *For the Learning of Mathematics, 32*(3), 15–21.

Roth, W.-M. (2013a). To event: Towards a post-constructivist approach to theorizing and researching curriculum as event*-in-the-making. *Curriculum Inquiry, 43*, 388–417.

Roth, W.-M. (2013b). Toward a post-constructivist ethics in/of teaching and learning. *Pedagogies: An International Journal, 8*, 103–125.

Roth, W.-M. (2013c). *What more? in/for science education: An ethnomethodological perspective*. Rotterdam, The Netherlands: Sense Publishers.

Roth, W.-M. (2014). Science language *Wanted Alive*: Through the dialectical/dialogical lens of Vygotsky and the Bakhtin circle. *Journal of Research in Science Teaching, 51*, .

Roth, W.-M., & Radford, L. (2010). Re/thinking the zone of proximal development (symmetrically). *Mind, Culture, and Activity, 17*, 299–307.

Roth, W.-M., & Radford, L. (2011). *A cultural-historical perspective on mathematics teaching and learning*. Rotterdam, The Netherlands: Sense Publishers.

Roth, W.-M., & Thom, J. (2009). The emergence of 3d geometry from children's (teacher-guided) classification tasks. *Journal of the Learning Sciences, 18*, 45–99.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge, UK: Cambridge University Press.

Star, S. L. (1991). Power, technology and the phenomenology of conventions; on being allergic to onions. In J. Law (Ed.), *A sociology of monsters: Essays on power, technology and domination* (pp. 26–56). London, UK: Routledge.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.). Cambridge, UK: Cambridge University Press.

Todd, S. (2003). *Learning from the other: Levinas, psychoanalysis and ethical possibilities in education*. Albany, NY: State University of New York Press.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57*, 69–85.

Vygotskij, L. S. (1934). *Myšlenie i reč': psixologičeskie issledovanija*. Moscow, USSR : Gosudarstvennoe social'noèskonomičeskoe isdatel'stvo.

Vygotskij, L. S. (2001). *Lekcii po pedologii* [Lectures on pedology]. Izhevsk, Russia: Udmurdskij University.

Making Decisions about Gambling: The Influence of Risk on Children's Arguments

Annie Savard
McGill University, Canada

Abstract : This article presents results from a study on decision-making towards eventual participation to gambling activities by grade 4 students. For this study, six learning situations were proposed in a fourth grade classroom. The researcher, who was also the teacher, proposed some extra activities in order to define gambling. Students learned about probability and developed, at the same time, the ability to think critically about gambling. She then proposed three fictional situations of gambling to the students, and asked them if and why they would (or would not) participate in the situation. By studying the arguments that students provided, she explored how students' probabilistic reasoning and critical thinking can influence their decision making towards these activities. This study therefore focuses on mathematical mobilization of resources in a citizenship approach. Results suggested that mathematical, sociocultural, and personal contexts were the determinants of which perspective the students were situated in. The results obtained show that the students' arguments were based on a variety of aspects: the affectivity towards a game situation, the ethical aspects of the proposed activity, the affective conceptions towards the random effect, or finally the mathematical aspects of the situation. Students were also able to talk about the risks they faced when participating in those gambling activities.

Keywords: decision making, gambling, risk, probability, critical thinking, mathematical context, citizenship competencies.

A Social Problem: The Risks Associated to Gambling

Gambling activities are part of the popular cultures in many societies (Korn & Shaffer, 1999). Gambling activities are activities where the outcome is mainly or totally based on chance and where there is an irreversible stake of money, object or action (Arseneault, Ladouceur, & Vitaro, 2001; Chevalier, Deguire, Gupta, & Deverensky, 2003). It is impossible to control or predict with certainty the outcome of the activity, even if some skills might be used in the activity. For instance, it is possible to make qualitative estimation of the chance of having the winning hands when playing poker. With the development of communication technologies, online gambling activities are now known worldwide, driven by the success of televised poker tournaments, online casinos and interactive lotteries (Chevalier & Allard, 2001). Risks involved in gambling activities are more often than not related to losing money. Gamblers lose money and then want to participate more to win their money back. This well-known phenomenon is referred to by psychologists is called "chasing losses". For some people, these particular activities can create a certain addiction that is sometimes hard to get out of and these addictions can have negative consequences on them and their surroundings (Ladouceur, 2000).

Some countries like Canada do not allow minors to gamble (Campbell, Derevensky, Meerkamper, & Cutajar, 2011; Chevalier & Allard, 2001). However, it seems that children do play these games (Campbell et al., 2011). In fact, children start participating in gambling games before the age of 12 (Griffiths & Wood, 2000; Gupta & Derevensky, 1998a; Ladouceur, Dube, & Bujold, 1994; Tremblay, Huffman, & Drabman, 1998; Wynne, Smith, & Jacobs, 1996). According to an important study conducted by Ladouceur, Dubé and Bujold (1994) with 1,320 elementary students from ages 8 to 12 living in the Quebec region, 86 % of the students had already won money and 37 % of them had even offered up one of their important possessions in a game. Some of these students had even won big amounts for their age, and more than 40% played gambling games at least once a week. Amongst these students, lottery-type games were the most popular, followed by bingo, card games, sports bets, specific event bets, video poker and finally slot machines. Works from Felsher et al. (2004) on the topic of lotteries reveal that the average age of children that play instant scratch games is 10 years old, 11 years old for lottery draws and 12 years old for sports lotteries. Not taking into consideration their degree of implication in those activities, sports lotteries is the most popular amongst children of all ages (Felsher, Derevensky, & Gupta, 2003). Lotteries form the first introduction to gambling games. Since the risks associated to lotteries are perceived by the general population to be negligible and that children do not seem themselves as vulnerable towards them, the popularity of lotteries remains very strong. Additionally, the promotion and positive publicity of these types of games seems to entice more people to engage in these games.

Gupta and Derevensky (1998b) argue that the status of the "social player" among teenagers, quickly becomes problematic player. According to the researchers, young people tend to develop a growing interest in these activities, which has the effect of increasing their participation and also the stakes. Unlike other types of addiction like smoking, alcohol or illicit substances, it is difficult to identify problem gambling among adolescents, because the manifestations of this addiction are not apparent (Lesieur & Klein, 1987). It is still possible to assess the degree of participation in these games by observing the adoption of high-risk behaviors, such as drinking alcohol, smoking or drugs (Proimos, DuRant, Dwyer Pierce, & Goodman, 1998). The more a teenager plays, the more likely (s)he is to manifest these behaviours. For example, some teenagers will instead use the money they have to purchase a meal for playing slot machines or other activities (Fisher, 1992; Griffiths, 1990; Hardoon, Deverensky, & Gupta, 2003; Wynne et al., 1996). The gambling habits of young problem gamblers push them to engage in risky and delinquent behaviours such as substance abuse, alcohol, theft and vandalism (Chevalier & Allard, 2001; Dickson et al., 2004; Fisher, 1992; Griffiths & Wood, 2000; Proimos et al., 1998. Winters et al., 1993. Wynne et al., 1996).

Teenaged gamblers have generally a negative attitude towards school (Wynne et al., 1996). These youth are so concerned with their gambling activities that they lose focus in attending and participating in lessons and being successful in exams (Derevensky & Gupta, 2001). Teenaged gamblers show a decreased interest in all components of academic life: studying, working , attending school, and developing positive relationships with school staff. Being late, absenteeism and delinquency are also common for these students. The consequences of these attitudes and behaviours are numerous: the declining school performance leading to failure, students go through sanctions: expulsion and even student dropout (Chevalier & Allard, 2001). Troubled youth are more likely to participate in gambling and deal with addiction because of their low grades, various hiccups from their personal history and the high dropout rate (Fortin, Ladouceur, Pelletier, & Ferland, 2001). Students with learning disabilities also show higher risks of becoming addicted (Marotta & Hynes, 2003). Low self-esteem is a common cause of adolescent gambling (Chevalier & Allard, 2001; Gupta & Derevensky, 1998a).

When gambling, youth and adults experience dissociative reactions. Gamblers experience cognitive illusion; they think they have a control on the gambling activity, which Langer (1975) named illusion of control. They think that external power, such as luck, faith or God might help them win. Usually, they don't perceive the risk of losing or being addicted because they believe in their good fortune. Additionally, they are detached from the reality around them because their cognitive functions are fully focused on the game (Gupta & Derevensky, 1998a; Ladouceur, 2004). In a gambling situation, the player loses track of time and feels another person. He feels out of himself and looks outside playing. In a trance, he loses his memory (blackout) of events that occur when playing (Jacobs, 2000; Rossen, 2001; Wynne et al., 1996). Among adolescents surveyed by Wynne et al. (1996), 75% of adolescents who were experiencing gambling problems reported losing track of time in a game situation and only 24% of those who were not experiencing gambling problems said they encounter this problem.

Teenager gambling is prevalent worldwide (Jacobs, 2000; Shaffer & Hall, 1996) and in many cases, it is perceived as an harmless pastime (Campbell et al., 2011). Teenagers gamble for fun, excitement, a challenge and for making money (Wynne et al., 1996). They gamble also for supporting charity, to pass the time with friends and out of curiosity. Thus, gambling may lead to positive emotions. Consequently, parents also do not seem to perceive the risks of gambling (Campbell et al., 2011). Yet, teenagers are at risk to develop an addiction to gambling. Teenaged gamblers do not see themselves as problem gamblers. Based on the work of Hardoon et al. (2003), teenagers minimize the severity of their problem. When they finally agree they need help, teenagers have serious family problems, social, academic and legal.

An addiction to gambling is hard to treat and thus, prevention is the best way to protect teenagers against risk (Ladouceur, Boudreault, Jacques, & Vitaro, 1999). This prevention should start before the habits of gambling begin to develop. Thus, it is important to provide students with the tools for gambling prevention as early as elementary school (Crites, 2003). Since children do not have the knowledge needed to evaluate the probability of winning as well as the risks of dependency associated to a particular gambling game, it is the school's responsibility to play its role and to give children the tools needed to do so. It is not enough to give students knowledge about the topic. A prevention movement about gambling games must be supported by the development of mathematical competencies and also citizenship competencies like critical thinking and decision-making.

The development of citizenship competencies as a way of prevention

Legendre (2004) reminds us that a competency is complex since it is the product of a dynamic organisation of its components. The mobilization of a person's resources represents more a competency than an accumulation of knowledge. It is by using and orchestrating its resources that a person demonstrates his competency (Perrenoud, 2002). The knowledge used is transformed through this process and it is thus an opportunity to expand the knowledge. Citizenship competency can be developed through the development of mathematical competencies. Ten Dam & Volman (2004) inform us that the goal of a citizenship competency is that each person as a member of a community of social practices participates in a critical and responsible way to the practice in question. Each member of a society participates in a democratic way to its evolution. It means that each member of the society has to be able to make choices and knowing why they are made, has to respect the choices and opinions of others, has to discuss the choices made, has to make his own opinion and has to share it with others (Halpern, 2003; Paul & Elder, 2001; Swartz & Perkins, 1990). Decisions can be made on the individual or collective level, depending of the needs of the person.

Making a decision implies consequences for oneself and others; there is an individual and social responsibility (Swartz & Perkins, 1990). Usually, the choice made tends to be oriented to instant gratification or pleasure, short-term effects (Paul & Elder, 2001). Considering long-term decisions requires a deeper thinking and self-discipline. Some consequences of long-term decisions can be known immediately or can be found later (Paul & Elder, 2001). According to Swartz & Perkins (1990) making a decision means generating and exploring different options and assessing them in order to find the best choice. Intuitively, one option usually emerges as the most desirable, but further consideration usually leads to select another option. Taking into consideration personal feelings is also part of the evaluation of the options. Identifying and recognizing needs are also part of the decision-making process, because they help to generate options. Halpern (2003) stated that this process takes in consideration personal characteristics such as values and knowledge, cognitive and cultural bias, and environmental variables (availability of an object for instance). Assessments reflect the "pros" and "cons" of the options and so rely on probabilities, possible consequences and possible risks or benefits involved. It comes to choosing the "best" option, but the best for whom? And by based on what criteria? Is the "best" option on a short or long term? These assessments should be critical to select an option that can meet the criteria identified by the person. Halpern (2003) raises an interesting point: the decision involves uncertainty about the outcome because we cannot know in advance the consequences of our actions. It is therefore important to be creative in the decision process to assess the maximum possible contingencies for our own wellbeing and that of others. Taking into consideration the context is also important, because the context is composed of socio and cultural factors, the kind of decision to make (for short or long term), technical issues or personal implications (importance of the decision). The process is dynamic and iterative: the nature of the decision, the context, or the options could be changed through it.

Critical thinking is intrinsically linked to decision-making process and contributes to identity development. Lipman (2003) believes that critical thinking has the function of facilitating the judgment based on criteria (including models), being self-correcting and takes into account environment. Criteria define the reliability and validity of a judgment; they are the basis of justification and comparison. Self-correction of critical thinking is by a self-assessment of thought, to correct our own mistakes. Taking into account environment or context involves to consider five aspects that can reflect an provide an overview of the circumstances and limitations of an event: 1) exceptional or unforeseen circumstances; 2) contingent or special constraints limits; 3) of an overview; 4) the possibility that the example is atypical; 5) the possibility that meaning is untranslatable from one context to another (Lipman, 2003). Taking into account the context allows one to determine the relevance of the criteria, to relativize the judgment and consider the implications. Affectivity is also part of the process to think critically, because attitudes influence thinking (Yinger, 1980). Personal values and biases have a strong influence on our cognitive structure and our implicit theories. Their influence colors every aspect of our thinking, but it is particularly strong in situations that require judgment or decision making about emotional or emotional issues (Yinger, 1980). Some attitudes are constitutive of critical thinking, such as respect for reasons produced (beliefs, actions and values) and respect for others and the intellectual authority. The latter is an attitude that is oriented towards research and inquiry, openness mind, the correctness of perspective, honesty and intellectual courage (Bailin, Case, Coombs, & Daniels, 1999). Note that critical thinking leaves room for doubt by giving up common ideas and in that sense, it is emancipatory. Creating a judgement or a point of view means also to be able to justify it. Explanation and justification are two different things (Duval, 1992-1993). Explanation allows production of reasons, while argumentation leads to examine the acceptability of these reasons during a justification. The acceptability of arguments is made based on two criteria: their strength and their relevance. The strength of an argument is based on two factors: the resistance against-one counter-argument and its epistemic value, defined as the degree of certainty or conviction tied to a proposal (Duval, 1991).

Critical thinking can be used in mathematics classes to examine, report and evaluate all aspects of a situation or problem, including collecting, organizing, storing, and analyzing information. Students need to be able to draw their own conclusions from the information and be able to identify inconsistencies and contradictions in the data (Krulik & Rudnick, 1999).

An Ethnomathematic Approach to Study the Probabilistic Structures

Inspired by Mukhopadhyay & Greer (2001), I developed an ethnomathematic model in order to construct learning situations with probabilistic structures (Vergnaud, 1990). This model contains three distinct contexts: sociocultural, citizenship and mathematical (Savard, 2008). The sociocultural context constitutes the starting point of the learning process by studying a particular object. This study can lead in to finding answers using mathematics. In that case, the mathematical context proposes a de-contextualisation of the object by means of mathematical "modelization". The results are then retaken to the original context in order to study the implications of the results on the object. The mathematics learned throughout the process can therefore serve as an assist to critical thinking and decision-making in the citizen context. The goal of this model is to take into consideration the complexity of an object or a phenomenon in order to make "sense" of what has been learned during the realisation of the process. In that sense, this model can also be used as a theoretical tool for implementing interdisciplinary approaches (Savard, 2011).

The learning of probabilistic structures has been studied under the angle of random events. In their work, Piaget and Inhelder (1974), Green (1988), Fischbein et al. (1991), Watson and Kelly (2004), Falk and Wilkening (1998) and also Jones *et al*., (1999) focused on studying students' comprehensions about certain random events. Other researchers have focused on the teaching of probability at the elementary school level (Brousseau, Brousseau, & Warfield, 2002; Medici & Vighi, 1996). In fact, the probabilistic structures can take 3 different forms: theoretical probability, empirical (or frequentist) probability and subjective probability (Caron, 2004; Konold, 1991). Some of these works were able to orient research towards student's probabilistic conceptions. In fact, Fischbein and Scharch (1997), Konold (1991), Brousseau (2005), Shaughessy, (1992), Amir and Williams (1999) and Kaheman and Tversky (1972) were able to identify certain conceptions that students had. Those conceptions are "explainable systems" that students give themselves in order to explain the result of an event. In this sense, they have their own domain of validity and we call them alternative conceptions (Savard, 2014). For example, students will explain the result obtained by throwing a 6-sided dice by the way it was thrown (Amir & Williams, 1999). This conception is called personalist interpretation. We have classified these conceptions according to the type of reasoning used: probabilistic or determinist (Savard, 2014). A probabilistic conception is based on probabilistic reasoning, where randomness and variability are key components. In the other hand, deterministic conceptions are based on the search of correlation between events. The deterministic reasoning employed does not allow a conceptual understanding of randomness and variability. The "conceptual complexification" (Larochelle & Désautels, 1992) of these conceptions, which is based on the process of learning by restructuring new knowledge within the existing one, therefore permits students to pursue their learning.

In this project, I tried using mathematics to answer a social problem: children gambling. I also tried to answer the following questions: How do grade four students perceive risk in gambling activities before and after learning about probability? What arguments are made during the decision-making process towards participating in a gambling activity?

A didactical experiment with elementary students

The didactical experiment lasted six months. Before hand, I designed a teaching experiment (McClain, 2002) composed of six learning situations. I used my ethnomathematic model to construct the learning situations that would highlight gambling situations, a development of mathematical competencies, and a citizenship development. Two learning situations were designed for each type of probability: subjective, theoretical and frequential (experimental). The contexts used in the two learning situations on subjective probability were on lucky charms and horoscope. Students were asked to do scientific experiments to test their lucky charms and their horoscope. Whole class discussions on their scientific experiments allowed them to express alternative conceptions, such as personalist interpretations or illusions of control in order to debate them explicitly. The contexts used in the two learning situations on theoretical probability focussed on dice and cards. Students were asked to do a bibliographic research on dice that included exploring a game board using two six-sided dice and exploring a deck of cards. Whole class discussions on their predictions allowed students to express some alternative conceptions about personalist interpretations and have them think about the meaning of the ratio (the number of favourable cases out of the number of all possible cases). The contexts used in the two learning situations on frequential probability were on spinners and coin tosses (Head or Tails). Students were asked to do experiments and record them. Whole class discussions on their results allowed them to think on the variability and uncertainty when addressing alternative conceptions, such as personalist interpretation. The researcher was also the teacher of this class of 27 fourth grade students (9 and 10 years old). The school was located in a suburban area of a larger city. All the teaching was done in French.

At the beginning of the didactical experiment (mid January), students were asked to draw their representations of gambling activities and then answer a questionnaire on their gambling habits and their knowledge on gambling and probability. Then, the six learning situations were implemented over three months. The students were given two formative assessments, one after third learning situation and the other after the last one. Those assessments were not graded nor reported as their learning performance (report cards). At the end of the teaching experiment, the students were asked again to draw their representations of gambling activities. The researcher interviewed each of them so they could explain the differences in their first and second drawings. Again, the same questionnaire was administered after the teaching experiment (mid April). I proposed three fictional vignettes at the very end of the didactical experiment two months after the students did their last drawings. Each vignette used a context that was closer to students' lives as much as possible. Those contexts were: a fund raising lottery, a coin toss game using personal objects, and betting on a race. The first two vignettes involved theoretical probability and the last one involved subjective probability. Each vignette was presented on different days. Students were first asked to respond individually on a sheet. The teacher/researcher collected them and facilitated a whole-class discussion on the vignette. After the discussion, students were again asked to write down what their answer to the question would be after the discussion on a new sheet.

Vignette 1: The Draw

A local youth group from your neighbourhood organize a draw to raise money for an organization. They sell 100 raffle tickets at $1 per ticket. A big prize of $25 is drawn. They asked you to buy a ticket with your money.

What do you do? Why?

How much profit will the local youth make?

Vignette 2: Coin Toss

Your friends propose to play a coin tossing game. Each participant puts an object that they like. The first of the three friends to get three faces in three tosses wins the objects.

What do you do? Why?

How many possibilities do you have to win the objects?

Vignette 3: The Bet

Your third neighbour thinks he or she is a race-running champion. But you know you can beat him or her because you are better than him or her. The neighbour challenges you to a race in the schoolyard. The first to arrive wins, and the loser has to give $2 to the winner.

What do you do? Why?

Comments:

All the discussions on the learning situations and the fictional vignettes were audio and video recorded and the content was then transcribed into text format. A pseudonym was given to every student in order to preserve the confidentiality of each participant. The Atlas.ti software was used to code the data. I used a mixed coding: some codes were determined and others emerged from the corpus. The data was then interpreted using Deblois (2003)'s cognitive activity interpretation model. This model tends to describe the dynamics involved in the student's comprehension in a school learning process. It takes into consideration the student's representations of the situation, the procedures used and the expectations provoked by the didactic contract. The coordination between these components can lead to a partial or general comprehension in other contexts. We also interpreted the data using my enthomathematics model. Each student's answers was situated according to the context they belong. For example, answers given by students on probability belonged to the mathematical context, where answers describing gambling activities or illusion of control belonged to the sociocultural context. In this article, I will present findings from the drawings and the interviews, the questionnaires and the three fictional vignettes.

Findings

I will present the students' perception of risk associated with gambling activities and the arguments they provided in fictional gambling situations.

Students' Representation of Gambling: Their Perception of Risk

In answer to the first research question that studied how grade 4 students perceived risks in gambling activities, our finding showed that at the beginning of the teaching experiment, very few students were able to recognize the risks involved in gambling activities. Six students were unable to represent gambling activities. In the first drawing, they represented game boards instead of gambling activities. Among the 21 students who represented gambling activities, six of them represented losing money as risks associated to gambling, and one of them wrote about gambling addiction. The responses given in the first questionnaire showed similar results. Thus, betting was perceived as a challenge or a harmless activity. Only 5 students were able to recognize the possibility of losing something. Wager was also considered as a challenging and harmless activity. It was also considered as a winning activity by many students. Only one student recognized the risk of losing money. The six questions on chance and luck showed that many students responded in terms of illusion of control. They did not respond in terms of uncertainty and randomness. They presented the same thinking than pathologic gamblers.

At the end of the teaching experiment after the enactment of the six learning situations, the students' representations of risk changed. In their last drawings, only three students as opposed to six still confused board games as gambling activities, and five students were able to make distinctions between games of chance and gambling. The risks associated with losing something were also represented. Students were able to talk about addiction because four of them talked about relatives who had an addiction in one of the learning situations. They said that they developed new knowledge towards those activities. Awareness to addiction is a fundamental aspect to preventing pathologic gambling; the assessment of probabilities of winning is another. In the second questionnaire, results suggested that betting did not seem to be problematic anymore because students were more cautious about the stake on question 1. Wagers seemed to still focus on winning on question 16. Some illusions of control were still present on the second questionnaire, but they were less frequent in all the questions asked.

The students admitted to participate in gambling activities. Many of them gambled, but it looked harmless to them because it was done between them.

Students' participation of gambling activities. In the questionnaires, two questions were asked to students in regards to their gambling habits. At first, they were not sure about their participation because many of them were confused with the expression: "I bet that". After reading carefully the questions, they saw that betting was linked with an object or money. The results of Question 2 (Did you already bet?***)*** were surprising. Ten students in the first questionnaire, and 14 in the second said they had gambled. Several reasons may explain this discrepancy. First, some students might have started to play for the first time in the interval between the two questionnaires. Second, students might have been able to recall having participated in a bet because of the context in the learning situations that served as a trigger. Third, students may have been more attentive to the types of activities that occupy their leisure time, since the learning situations led them to think about different types of activities. I believe that the learning situations were not an incentive to play, but that possibility existed nonetheless.

The answers to Question 3 (If you've ever participated in any of these activities, circle which one or ones: Marbles [the winner kept the marbles]; Lottery tickets [6/49]; Dice when betting something [ex: an object or money]; Cards when betting money; Spinner games betting money [ex: fair]; Scratch tickets like seen on tv [ex : popular tv show name]; Betting an object [ex: cards, figurines, …]; Betting an action [ex: a service,…]; Betting money; Participating at a drawing by buying a ticket; Head or tail when betting something; None of theses answers) of the two questionnaires show that the game of Heads or Tails seemed the most popular. However, students used this activity to determine the winner of an object or to determine which player would play first rather than bet. It was a way to make decision instead of a challenge. Some activities such as lotteries are prohibited for children under 18, but students said they have participated in these activities.

Visual representations and interviews. The first drawings made by the students showed that 21 of them knew some gambling activities. The first 27 drawings were classified based on students' representations. Four categories appeared in the light of the representations. The first category was not a gambling activity; it was related to board games.

1. Board games: 6
2. Casino activities: 11
3. People who gamble: 4
4. The emotions of people who gamble: 6

Popular activities in the casino (slot machine, roulette and blackjack) were represented. Four students associated the number 7 to casino operations.

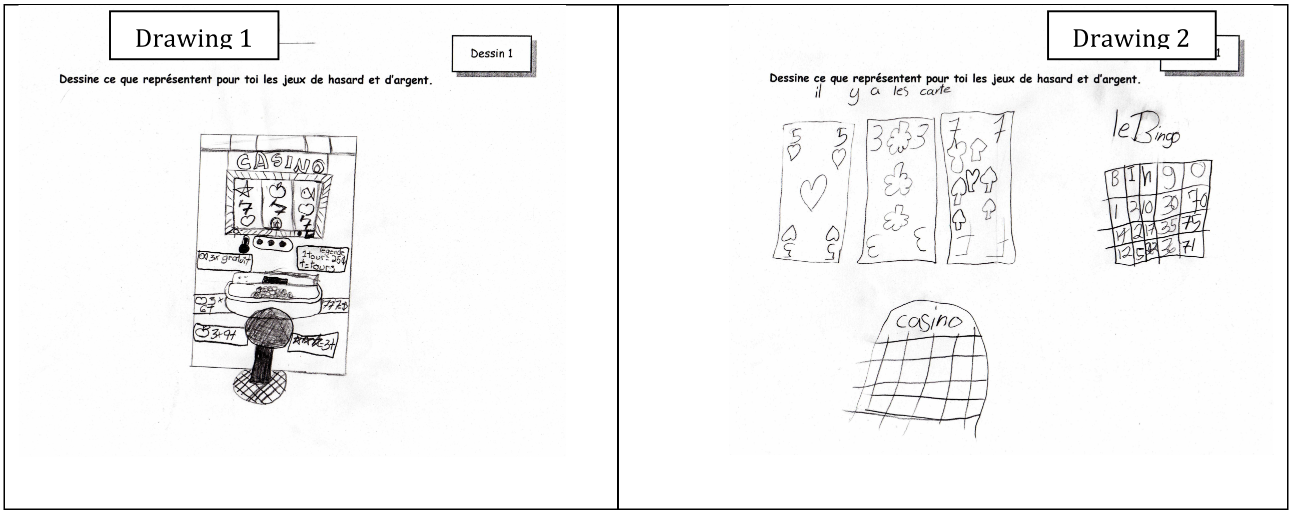


Figure 1. Students' representations of gambling

The visual representations drawn by the students show that these adult activities were familiar to them. However, there appeared to be confusion between such activities and board games such as Monopoly. Indeed, the money from these board games was an element of confusion for some students. They wondered about it aloud when performing the activity of drawing. In the drawings, students also represented adult characters playing. Six students were positioned in a citizenship context: they chose to represent the emotions of people who gamble. The characters expressed emotions and text in the drawings helped understand them. Unlike the other drawings, these drawings were dynamic: they were drawn like a cartoon. The characters spoke or explanatory text described the representations and in general, the drawings represented a moral (four drawings). This moral vehicled the following message: gambling a lot leads to losing money. The participants seemed to express an addiction to many gambling activities where people bet their money earned and they stop when they lost everything. One student even highlighted the savings to show the money obtained. This student expressed being against making money easily and losing all easily too. It is possible that an implied contract could have be established between the teacher and students. This contract would have shared implicit expectations with teaching contract as it was presented. Thus, this implicit contract could have led students to believe that this type of activity is not acceptable:

In the second drawings, students' representations were different. Although the first category was not related to gambling, it was still present. The three categories were changed to reflect the evolution of the students' representations:

1. Board games: 3
2. Gambling activities: 14
3. Distinction between games of chance and gambling: 5
4. Gambling as a risky activity: 5

Changes in the categorization of the final drawings showed that students were more familiar with gambling activities. Indeed, the categories were refined since a category on the distinction between games and gambling emerged, and one of them changed. The category on emotion changed to a category called gambling as a risky activity. Four drawings represented gambling as losing money and one drawing represented an addiction to gambling.

The interviews with each student revealed that some students still confused board games and gambling activities. Board games do not cause addiction, unlike gambling activities. The risks of addiction was also one of the areas identified by some students during the interview. The scope of participation, the monetary loss and the hardness to stop playing was cited as the information learned in the learning situations. Students talked about some personal experiences with gambling when discussing their drawings:

Annie: Hello Magalie, can you explain your drawing?

Magalie: I have done a drawing that represents the casino and then for me what that means is the casino is just wasting your money. Here I did: I lost $ 1,000 I lost $ 10,000 and $ 2,000.

Annie: And why is it losing and people will go to the casino to win?

Magalie: because they say: oh I will bet on so many businesses, as I think this time if it will be the lucky one, while I think it's really lost. I say that because my grandmother owes a lot of things to the casino.

Annie: She owes a lot of things? What do you mean?

Magalie: She owes a lot of money to just anyone.

Annie: Ok, she borrows money to gamble. Your own grandma?

Magalie: Yes.

Annie: So gambling and you have drawn a casino.

Magalie: Yes. Because that is what makes me think of gambling.

Annie: So everything we see here is people who lose. Is that right?

Magalie: Yes.

Four students mentioned knowing relatives who have an addiction. On the other hand, some students qualitatively assessed the probabilities of gambling situations. Their assessments were supported with correct mathematical knowledge. I believe that these students now had some knowledge allowing them to evaluate other gaming situations. In this regard, a student showed the use of a quantification of probabilities to encourage gambling:

Annie: Where are these people?

Sarah: In the casino.

Annie: Okay.

Sarah: And then they say they can make money. Then I forgot to write probabilities ...

Annie: You wanted to write the probabilities?

Sarah: Yeah, I wanted to write one chance out of ...

Annie: Why probabilities at the casino?

Sarah: Well, because it can encourage people, they say maybe it would be the fiftieth ...

Annie: That would be the fiftieth? What do you mean?

Sarah: Well if it's a 50 chance of winning maybe people will say it's the fiftieth time they play and they have not won. Maybe I'll win (E.671).

This student adopted the posture of a casino tenant to entice people to play instead of the posture of a gambler.

The changes in the categorization for the final drawings show that students were more familiar with gambling activities from a sociocultural context. The categories were refined since a category on the distinction between games of chance and gambling emerged, and one of them had changed. A category on emotions when gambling became an emotion-focused in a loss, also showing the possible addiction. The five students who made those drawings were located in the citizenship context.

The questionnaires. In the questionnaires, nine questions were asked on gambling activities and luck. The results of the first survey Question 1(What do you think it means to bet?) showed that the bet is more understood as a challenge or as a harmless activity. The emotional aspect was also present in the material presented, since you could lose or win. The answers to the second survey showed that the challenge of the activity became an issue that involved a bet that you could lose or win (9 responses more). Students thus abandoned the emotional perspective of the activity from the knowledge of its function. If students are more interested in how the activity works, it is possible that eventually they could end up being able to quantify the odds of winning in these activities.

Questions 5, 6, 8,10 and 12 were related to the affective conception named illusion of control. Some explanations of Question 5 (What are your chances of winning if you choose your lottery numbers? Circle the answer: Good - Better - Same - Worse – Bad. Explain.) showed that the control of chance is possible: the fact of choosing numbers affected the odds of winning. In the second questionnaire, six students have however shown the "complexification" of their affective conceptions and felt that the possibilities were the same, picking the numbers or not. The answers to question 6 (Can a lucky charm help you to win? Explain.) showed that the majority of students (14 out of 27) did not believe that luck had an effect on the chances (Q1 and Q2). According to them, a good luck charm could influence the location of a player by giving him confidence and therefore increasing the possibilities of winning. The explanations by students to question 8 (If your horoscope predicts you are lucky, are your chances of winningy: Good - Better - Same - Worse- Bad? Explain.) were divided according to if they believed or not in horoscopes. Among those who didn't believe in them, some explanation still showed that they might win. Perhaps students rated the situation from a probabilistic point of view, since it is possible that the event occurs. Note that in the second questionnaire, the responses to this question showed a clear tendency to question the effectiveness of horoscopes. Several events of illusion of control were revealed in questions 10 and 12 in the two questionnaires. Question 10 (Is there a number that is more lucky than others? Which one et why?) showed that students attributed certain famous figures as lucky. The students' explanations to question 12 (In the game of the lucky wheel, are your chances of winning better if it is you who spin your wheel? Why?) showed few ideas related to the handling of deterministic conception. It was more the person making a difference when spinning a lucky wheel and not the action. These deterministic conceptions seemed to be in the complexification process when responding to the second questionnaire. Besides the fact that the students' responses were less based on deterministic conceptions, the explanations were more based on mathematical learning acquired in the classroom.

In the first questionnaire, the students' explanations to question 16 (What does waging mean?) indicated that the action of wagering drived from affectivity in response to the challenges. Positive aspects about winning were mentioned by 12 students, but nothing about losing. One student said that it was wasting money. In the second questionnaire, three students mentioned the effects of chance in the context. These students showed that they might have considered chance in another context view. This category was not present in the initial questionnaire. However, seven students explained chance by giving answers coming from other contexts of games of chance and gambling. The questions students have were linked, in both questionnaires, to gambling activities more than just prediction or chance.

Students' decision-making process about gambling activities

In answer to the second research question that studied the arguments deployed during the process of making a decision on an eventual participation to gambling activities, the results showed that the arguments were based on the affectivity manifested towards a gambling situation, on the ethical aspects of the proposed activity, on the alternative conceptions towards the random effect, and finally on the mathematical aspects of the situation.

Affectivity. The affectivity manifested towards a gambling situation had aspects related to the impact of the activity: the gains or losses from the activity or the challenge. For example, the vignette on the coin toss game showed that the emotions related to the choice of the object was the main argument to decide whether to participate or not because the students said they did not want to lose it. These students stressed emotions associated with the object by the consequences of its loss. Thus, they would be saddened or their friends would be sorry to have lost. The loss outweighed the possible gains. These students evaluated a greater risk. In the case of the vignette on the bet, some students explained that some players could participate more in order to recover the money lost in betting. Early in the discussion following the writing assignment, the value given to the objects was the key point to decide. This was the possible loss of the object that override the possibility of winning. In some case, the gain didn't seem to be highly considered, the emotional value predominated. Decision-making was based here on the possible gain or loss of the object and not on the mathematical possibilities of winning. Besides, if the issue was not an important object, it became easier to participate, because the gambling would have a positive impact: it would be a winner and the other friends would have no trouble losing their object. Yet it was the opposite argument used by another student in the case of an item worthless to him because he cared about the feelings of his friends who might lose if he won. He based his decision on the emotional aspect of the result: two losers and a winner. This emotional aspect was actually an ethical argument against the consequences of the game. One student said she didn't like this type of game and that's why she said no. She did not provide other reasons. Knowing the student and her family, I think that her religious beliefs disapprove gambling activities.

Early in the discussion on betting for a race, many students expressed their rationale for participating or not. The reasons were based on losing or winning money and on the dangers of the race. For losing or winning money, some students explained that their decision was based on the amount involved because they did not want to lose large amounts, but would participate if it was a small amount. As for the dangers of the race, the discussion targeted two reasons: the danger of being injured and of becoming addicted to gambling. The danger of being injured could be caused by falling, getting hit by a car or hustled in the race. Winning money, having fun or trying to win back the money lost were the dangers associated to being addicted to gambling. In other cases, as in the vignette of the draw, students said they wanted to earn the amount of $25. Some were looking for a monetary gain. The money at stake was the element considered by the students to decide whether to participate or not. Six students chose to participate with hopes of winning while three chose not to participate because of the risk of losing. One of these students based his thinking on mathematical aspects to assess its low probabilities of winning while another evaluated the small amount of profit.

Affectivity was also linked to the activity itself. Students based their decision on the assessment of this activity: like tossing a coin or not liking betting. Others wanted to take up the challenge proposed by the race. Provocations linked to betting could encourage them to participate. Some students wanted to prove that they were better runners than the neighbour. The money was not taken into consideration.

Some students said that they would have run without the money, because racing was fun or because they didn't like betting. Two students added they would not participate because of their religious beliefs. Students who based their answers on this aspect were located in the citizen context.

Ethical aspect. The ethical aspect of the proposed activity could be declined under three aspects: the purpose of the activity, the impact of the activity, and the risks involved during the activity. The goals of the activity in the vignette on the draw related to the money raised. Students based their decisions on the support they would make to a charity. For 16 students, the support provided to an organization was the central argument that influenced their decision. They based their decision on the ethics of the situation instead of their personal comfort. Some even said that it was a gift and that the gain was not important for them. They chose the ethical aspect of the context as a criterion for deciding whether to participate or not. This instance highlighted how the purposed of the activity influenced their decisions. Others said that they did not know where the profits raised would go, and it was the central argument in their decision-making. The two students who questioned the integrity of the organization questioned the validity of their ethical action, as they assessed the charity organism before deciding if the ethics was important enough to invest money. They also took into account the context and were given criteria to decide. It showed the impact of the activity. For instance, in the discussion following answering the vignette, a student raised the legality of buying lottery tickets:

> *Marco:* A lottery, is all the time 18 years old and over. Well, we were not supposed to buy it, that I would say no there (V1.202).

The student questioned his participation from a legal point of view. While the discussion addressed the ethical aspects of the money raised, it addressed a critical aspect in a decision-making: evaluating options from a legal point of view. This awareness led to a student to consider saving money rather than to enrich himself, slippery wealth sharing to increase his personal wealth. The discussion moved them to personal wealth.

Students reported that there were risks in gambling activities,, such as having a physical injury during a race or hurting others. Students also mentioned the risks of addiction to gambling and money. Students who based their answers on this aspect were located in the citizen context.

Alternative conceptions. Alternative conceptions of chance involved the illusion of control or religious beliefs. In the vignette Heads or Tails, for example, two students relied on affective conceptions on chance. One considered that the face side of the coin was luckier while another considered himself unlucky: she wrote that she doesn't have much luck. In the discussions, a student said that he wanted to cheat when tossing the coin. He wanted to participate, since he believed he could control the outcome of the game. He showed a deterministic conception of the personalist interpretation. He thought that he could control the outcomes.

During the discussion on the vignette on the bet, two students based their decision on their religious beliefs that prohibited betting. Students who based their answers on this aspect were located in sociocultural context. Students who did express alternative conceptions did not mention any risk associated to the activity.

Mathematical aspects. The mathematical aspects of the situations were related to the theoretical and subjective probabilities of winning. Students who based their answers on these aspects were located in the mathematical context. For instance, the proposed context on the draw was a great influence on the decision-making process since the arguments used by the students were mainly based on the support given to the charity rather than participating in gambling activities. When the sociocultural context has shifted to a mathematical context, students considered the money invested, the money involved, and the money raised. The arguments used then were based on mathematical facts:

> *Mia:* Well, I say no because even if it's just for $1, then, well you've don't really have too much chances of winning if you buy a ticket. You have one chance in 100 of winning (V1.182).

Another student compared the possible benefits of buying a ticket. She addressed the expected gain. She based her rationale on the quantification of the gain and not on the probability of winning:

> *Magalie:* I would say n, because first, they will have the triple in their pockets. Then, he said that just $1, but maybe you have ... basically, is that you do not earn $25. You win, you win $24. That is like a catch because you spent $1 and won $25. At the end, you won $24 (V1. 182).

She estimated that her probabilities of winning were low. She took into account that all the tickets would be sold to show an awareness of the concept of the relative theoretical probabilistic structures, or 1 chance in 100. In fact, where the money went was more important than the money to be won. In fact, it was the only risk involved given the fact that the majority of students did not care about winning or losing money in this case. The citizenship context used all the space and obscured the mathematical context.

In the vignette on Heads or Tails, two students relied on mathematical aspects, specifically on the equiprobability of the outcomes. In the discussion, a student thought that all friends had the same chances to win: I chance out of 3. He based his decision on the mathematical aspects of the situation. He based is quantification on the number of participants rather than the number of throws to assess probabilities. However, it was fair to say that the participants had the same chance to get face-face-face or 1/8. He was not the only student that made this confusion. Another student was not able to think about the number of possible cases because she thought about counting all the attempts made by the participants. The context of gambling dominated. In fact, the gambling situation took all the space because another student did not see the need to focus on the mathematical context since in any case, she did not want to risk losing an object and had no chance of winning:

> *Mia:* That's because I, if I'm not involved, I cannot really have a chance to win (V2.282).

She did not show any awareness that the count of her probability to win could have helped her decide whether to participate or not. Only the assessment of the object led her to decide. On the other hand, another student was not able to count them because it was not possible to predict the next outcome anyway. That student confused the enumeration of probabilities and the outcome of the gambling activity. Doing so, it was impossible to use a mathematical argument to support his decision.

The teacher then asked the students to individually answer the question on a blank sheet. The study of the students' productions showed that 7 students counted the throws. Four students answered 1/8, when a student forgot some cases and answered 1/6. A student counted eight possibilities but said 3/7 because "you cannot have just face-face-face because everyone would win all the time." A student counted the combinations instead of the permutations and replied ¼. Some students took into consideration the players and the throws: six students answered 3/9, seven students answered 1/3, and one student answered 3/3. Four students used a deterministic reasoning by saying that it was chance. A student showed a conception of personalist interpretation because he said that it depended on the player,

In the case of the vignette on the race bet, two students relied on the information available to them to assess the situation. They evaluated qualitatively the subjective probabilities to win:

Melissa: For me, it depends. If I know I will win, then I may be saying yes, and then if I know I can lose, or you know I can lose my $2, then I will say no. You know, what I mean by that is, I might already have done a race against that person then I won ... maybe (V3. 37).

Martin: If ... it was not running but cycling, it would not have been hard, I would have immediately said yes because ... I think I would be able to beat him (V3. 319).

Students did not often use the use of probability to justify a decision.

Discussion

This study led to the emergence of four themes that identified the arguments used by elementary school students to support their decision of whether to participate or not in a gambling activity. Some students linked their reasons toward the risks to lose money or object, being injured, making sad their friends or being addicted to gambling. In fact, the risks were involved into three themes: the affectivity manifested, the ethical and the mathematical aspects. No risk was mentioned in the alternative conceptions theme.

Wynne et al.'s (1996) work allowed me to compare the reasons that explain the participation of different teenager gamblers: teenagers without problem, at-risk or pathological gamblers. The motivations raised by adolescents related to the emotions shown towards the gambling situation, the ethical aspect of the proposed activity and distractions. The results from this study suggest that elementary school students also justify their participation by the emotions shown towards the gambling situation and the ethical aspects of the proposed activity. However, unlike the teenagers surveyed by Wynne et al. (1996), the students in this study also justified their eventual participation by alternatives conceptions toward chances and by the mathematical aspects involved in the situations. I hypothesized that the learning situations enacted in this study would influence students' responses.

The four themes that emerged from the decision-making process by students can be categorized according to a predominant context. Thus, the alternative conceptions towards the random effect were elements that belonged to the sociocultural context. The mathematical aspects of the situation were elements that belonged to the mathematical context. The affectivity manifested towards a gambling situation and the ethical aspects of the proposed activity were elements that belonged to the citizenship context. The risks perceived by students belonged to the mathematical context or the citizenship context. It is interesting to note that those contexts might used to prevent gambling addictions.

The theme affectivity manifested towards a gambling situation, emerged more often from the students than the other themes. When students used arguments belonging to the this theme, they situated themselves in the citizenship context. The themes mathematical aspects of the situation, and ethical aspects of the proposed activity were not used often by the students to make a decision. It does not mean that they didn't consider them at all. They might have considered elements from the mathematical and the citizenship contexts, but they were not selected as key points to make a final decision. The decisions taken by the students did not all seem to have followed the process of decision making from the theoretical framework. Indeed, students reported that their religion forbade them to bet. In doing so, they did not seem to have generated more options to decide or re-evaluated their decision. I did not have access to the whole process to generate and evaluate all the options. Even if it was possible to grasp some options that were considered by the students throughout the whole class discussions, I did not have enough material to have a strong interpretation. However, I can say that the students used the context of the gambling situation as a central element of the decision-making process. The same framework had not only served to generate options, but was also used to justify them.

Another limit of this study was related to the students' perceptions of risk. Throughout the whole didactical experiment, the students were not asked directly about how they perceived risk. Instead, I wanted to know if it was a part of their general understanding of gambling activities. I didn't want to be bias or make judgement on their relatives who gambled. It could be perceived as a limitation of this study at this point, but the opposite could have affected the complexification process of their alternative conceptions, especially the ones related to the illusion of control. When we talked about those, some students mentioned how their relatives used lucky charms. If, as a teacher I presented those thinking negatively, it might have had created an affective conflict with students. Then, it would have been harder for them to think about those ideas from a mathematical point of view.

The vignettes proposed presented some limitations. Thus, the contexts proposed in the vignettes were inaccurate for some students. The context proposed in the first vignette was raise-funded tickets lottery. The number of tickets sold was not given to students; they had to assume that all of them would be sold. Students could not assess the amount of prizes because they represented a situation in which few tickets had been sold. In addition, determining the profit for the winner and for the organizer of the draw was a challenge for 15 students. Does it mean that understanding the distribution of money had an effect on the decision-making process? That could be a factor of some importance, since the fact that the profits of the draw would go to a charity organism had an impact. In fact, the choice of the charity organism was found to be an important aspect of the situation. Students evaluated the seriousness of the organizers of the draw instead of doing a mathematical evaluation. After the discussion, the students answered the second time the questions on the first vignette. The teacher added some information to the context that helped students to define it. Thus, the charity organism in question was a known and reliable one and all tickets were sold. These clarifications were required to discuss the same context for all.

The context presented in the second vignette asked to quantify the probability of a compound event when playing Heads or Tails, which students were not able to do. It was possible to make an experiment, since they had access to manipulatives. However, no student used the material to make the experiment and thus answered questions. Despite the fact that students have counted the results of throwing a coin, none reached a solution that showed an understanding of enumeration. Although the proposed location could be addressed from a theoretical approach, the complexity of the rigorous theoretical analysis of this situation might have allowed, at best, to calculate the probability of winning for one trial and use equiprobability to predict the probability of winning. During the whole-class discussion, the teacher asked a question about the mathematical background and she noticed that the students used the number of participants rather than the number of throws. She then wrote the following question on the board: What is the probability of getting face-face-face when tossing the coin three consecutive times? She explained the issue and asked the students to answer on a sheet. After picking up the students' sheets, the discussion continued on the procedures used to count the results. The teacher constructed with the students a tree-diagram to create the sample space. The calculation of a compound event seemed to require a representation of the situation that was challenging for fourth grade students.

The nature of the objects involved in this vignette was also very important. The students discussed more the objects than the probability to win. Their arguments were on the risks of losing a valuable item, the taste of winning objects, and making friends sad if you win. The probability to win and alternative conceptions were also arguments to decide in some cases. Again, the sociocultural context took over and left little room for mathematics in the situation. The main arguments were based on the outcome of the game and the risk of losing a valued object. Very few students used mathematical arguments.

The context presented in the third vignette was taken from school. The school environment brought students to visualize races conducted as part of the physical education classes. Some of the risks raised by students were those they faced when racing in the schoolyard: being injured or hurt by another student or a car (the street was pretty closed to the schoolyard). They also focused on the running skills of each participant, which was part of the information needed to construct the subjective probability. In this case, some students used a mathematical context to decide.

Despite these limitations, the vignettes were fertile fictional contexts. The discussions yielded important information about the context of gambling activities. They were important to clarify the contexts presented. It was also a nice opportunity for students to revise their initial thoughts. Interesting enough, after the discussion on the vignette Heads or Tails, two students made a decision based of the illusion of control, which they did not do that first. It seemed that those conceptions were hard to complexify and can live in parallel with other reasoning. In non-fictional context, much information must be processed before deciding and some of them require significant cognitive work. That's why students have to situate themselves in different contexts: citizen, sociocultural and mathematics to make a decision. Since the mathematical context is embedded within the two other contexts, it may be more difficult to reach it.

Conclusion

This study shed light on students' perceptions of risk in the case of gambling activities. Some students did participate at gambling activities by betting, money, action or object. Like Ladouceur et al. (1994) pointed out, they still have access to adult lottery such as scratch tickets. Prevention should be a part of the curriculum. Like drugs, cigarette and alcohol are discussed over the years in the schooling system, risk associated to gambling should be address when teaching probability. It is absolutely necessary to address it because even if the risks are not explicitly addressed, students might have alternative conceptions such as personalist interpretation and illusion of control when learning probability. These conceptions do not help assess the risks in a mathematical way. Those conceptions can live in parallel in students' mind and create an obstacle for developing a conceptual understanding of probability. Thus, supporting students to complexify their alternative conceptions should be part of the learning conditions created to develop probabilistic reasoning.

Acknowledgements

I want to acknowledge the support grant from the Fonds de recherche société et culture du Québec (FQRSC) and thank my colleagues Limin Jao and Marta Kobiela for their incredible support.

References

Amir, G.S., & Williams, J.S. (1999). Cultural Influences on Children's Probabilistic Thinking. *Journal of Mathematical Behavior, 18*(1), 85-107.

Arseneault, L., Ladouceur, R., & Vitaro, F. (2001). Jeu de hasard et consommation de substances psychotropes: Prevalence, coexistence et consequences. / Gambling and Consumption of Psychotropic Drugs: Prevalence, Coexistence and Consequences. *Canadian Psychology, 42*(3), 173-184.

Bailin, S., Case, R., Coombs, J.R., & Daniels, L.B. (1999). Conceptualizing Critical Thinking. *Journal of Curriculum Studies, 31*(3), 285-302.

Brousseau, G. (2005). Situations fondamentales et processus génétiques de la statistique. In A. Mercier & C. Margolinas (Eds.), *Balises pour la Didactique des Mathématiques* (Vol. 165-194). Grenoble: La pensée sauvage.

Brousseau, G., Brousseau, N., & Warfield, V. (2002). An Experiment on the Teaching of Statistics and Probability. *Journal of Mathematical Behavior, 20*(3), 363-411.

Campbell, C., Derevensky, J., Meerkamper, E., & Cutajar, J. (2011). Parents'perceptions of adolescent gambling: A Canadian national study. *Journal of Gambling Issues*(25), 36-53.

Caron, F. (2004). *Splendeurs et misères de l'enseignement des probabilités au primaire.* Paper presented at the Colloque du Groupe des didacticiens des mathématiques du Québec (GDM) 2002: Continuités et ruptures entre les mathématiques enseignées au primaire et au secondaire, Université du Québec à Trois-Rivières.

Chevalier, S., & Allard, D. (2001). Pour une perspective de santé publique des jeux de hasard et d'argent (pp. 53): Institut national de santé publique du Québec.

Chevalier, S., Deguire, A.-L., Gupta, R., & Deverensky, J.L. (2003). Jeux de hasard et d'argent *Enquête québécoise sur le tabagisme chez les élèves du secondaire (2002). Où en sont les jeunes face au tabac, à l'alcool, aux drogues et au jeu?* (pp. 175-203): Institut de la statistique du Québec.

Crites, T.W. (2003). What Are my Chance? Using Probability and Number Sense to Educate Teens About the Mathematical Risks of Gambling. In H. J. Shaffer, M. N. Hall, J. Vander Bilt & E. George (Eds.), *Futures at Stake: Youth, Gambling, and Society* (pp. 63-83). Reno/ Las Vegas: University of Nevada Press.

Derevensky, J.L., & Gupta, R. (2001). Le problème de jeu touche aussi les jeunes. *Psychologie Québec, 18*(6), 23-27.

Duval, R. (1991). Structure du raisonnement déductif et apprentissage de la démonstration. *Educational Studies in Mathematics, 22*, 233-261.

Duval, R. (1992-1993). Argumenter, démontrer, expliquer: continuité ou rupture cognitive? *Petit X, 31*, 37-61.

Falk, R., & Wilkening, F. (1998). Children's Construction of Fair Chances: Adjusting Probabilities. *Developmental Psychology, 34*(6), 1340-1357.

Felsher, J.R., Derevensky, J.L., & Gupta, R. (2003). Parental Influences and Social Modelling of Youth Lottery Participation. *Journal of Community and Applied Social Psychology, 13*, 361-377.

Fischbein, E., Sainati Nello, M., & Sciolis Marino, M. (1991). Factors Affecting Probabilistics Judgements in Children and Adolescents. *Educational Studies in Mathematics, 22*, 523-549.

Fischbein, E., & Schnarch, D. (1997). The Evolution with Age of Probabilistic, Intuitively Based Misconceptions. *Journal of Research in Mathematics Education, 28*(1), 96-105.

Fisher, S. (1992). Measuring Pathological Gambling in Children: The Case of Fruit Machines in U.K. *Journal of Gambling Studies, 8*(3), 263-285.

Fortin, J.M., Ladouceur, R., Pelletier, A., & Ferland, F. (2001). Les jeux de hasard et d'argent chez les adolescents et adolescentes en difficulté. *Revue canadienne de santé mentale communautaire, 20*(1), 135-151.

Green, D. (1988). Children's Understanding of Randomness: Report of a Survey of 1600 Children Aged 7-11 Years. In R. Davidson & J. Swift (Eds.), *The Proceedings of the Second International Conference on Teaching Statistics* (pp. 287-291). Victoria, B.C.: University of Victoria.

Griffiths, M.D. (1990). The Acquisition, Development, and Maintenance of Fruit Machine Gambling in Adolescence. *Journal of Gambling Studies, 6*(3), 193-204.

Griffiths, M.D., & Wood, R.T.A. (2000). Risk Factors in Adolescence: The Case of Gambling, Videogame Playing, and the Internet. *Journal of Gambling Studies, 16*(2/3), 199-225.

Gupta, R., & Derevensky, J.L. (1998a). Adolescent Gambling Behavior: A Prevalence Study and Examination of the Correlates Associated with Problem Gambling. *Journal of Gambling Studies, 14*(4), 319- 345.

Gupta, R., & Derevensky, J.L. (1998b). An Empirical Examination of Jacobs's *General Theory of Addictions*: Do Adolescent Gamblers Fit the Theory? *14, 1*, 17-49.

Halpern, D.F. (2003). *Thought and Knowledge: An Introduction to Critical Thinking (Fourth Edition)*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hardoon, K., Deverensky, J.L., & Gupta, R. (2003). Empirical Measures vs. Perceived Gambling Severity Among Youth; Why Adolescent Problem Gamblers Fail to Seek Treatment? *Addictive Behaviors, 28*, 933-946.

Jacobs, D.F. (2000). Juvenile Gambling in North America: An Analysis of Long Term Trends and Future Prospects. *Journal of Gambling Studies, 16*(2/3), 119-152.

Jones, G.A., Thornton, C.A., Langrall, C.W., & Tarr, J.E. (1999). Understanding Students' Probabilistic Reasoning *Developing Mathematical Reasoning in Grades K-12* (Vol. 1999 Yearbook). Reston, Virginia: National Council of Teachers of Mathematics.

Kaheman, D., & Tversky, A. (1972). Subjective Probability: A Judgement of Representativeness. *Cognitive Psychology, 3*, 439-454.

Konold, C. (1991). Understanding Student's Beliefs about Probability. In E. V. Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education* (pp. 139-156). Dordrecht: Kluwer academic publishers.

Korn, D.A., & Shaffer, H.J. (1999). Gambling and the Health of the Public: Adopting a Public Health Perspective. *Journal of Gambling Studies, 15*(4), 289-365.

Krulik, S., & Rudnick, J.A. (1999). Innovative Tasks to Improve Critical- and Creative-Thinking Skills *Developping mathematical reasoning in grades K-12* (Vol. 1999 Yearbook). Reston, Virginia: National Council of Teachers of Mathematics.

Ladouceur, R. (2000). *Le jeu excessif : comprendre et vaincre le gambling*. Montréal: Éditions de l'Homme.

Ladouceur, R. (2004). Perceptions Among Pathological and Nonpathological Gamblers. *Addictive Behaviors, 29*, 555-565.

Ladouceur, R., Boudreault, N., Jacques, C., & Vitaro, F. (1999). Pathological Gambling and Related Problems Among Adolescents. *Journal of Child and Adolescent Substance Abuse, 8*(4), 55-68.

Ladouceur, R., Dube, D., & Bujold, A. (1994). Gambling Among Primary School Students. *Journal of Gambling Studies, 10*(4), 363-370.

Langer, E.J. (1975). The Illusion of Control. *Journal of Personality and Social Psychology, 32*(2), 311-328.

Larochelle, M., & Désautels, J. (1992). *Autour de l'idée de science : itinéraires cognitifs d'étudiants et d'étudiantes*. Sainte-Foy: Presses de l'Université Laval ;.

Legendre, M.-F. (2004). Cognitivisme et socioconstructivisme. Des fondements théoriques à leur utilisation dans l'élaboration de la mise en oeuvre du nouveau programme de formation. In P. Jonnaert & A. M'batika (Eds.), *Les réformes curriculaires. Regards croisés* (pp. 13-47). Sainte-Foy, Québec: Presses de l'Université du Québec.

Lesieur, H.R., & Klein, R. (1987). Pathological Gambling Among High School Students. *Addictive Behaviors, 12*, 129-135.

Lipman, M. (2003). *Thinking in Education (Second Edition)*. New York: Cambridge University Press.

Marotta, J.J., & Hynes, J. (2003). Problem Gambling Prevention Resource Guide for Prevention Professionals (pp. 62). Salem, OR: Oregon Departement of human Services, Office of mental health and addiction services.

McClain, K. (2002). A Methodology of Classroom Teaching Experiments. In S. Goodchild & L. English (Eds.), *Researching Mathematics Classrooms: A Critical Examination of Methodology* (pp. 91-118). Wesport, Connecticut: Praeger.

Medici, D., & Vighi, P. (1996). Une histoire... improbable. Introduction des probabilités à l'école primaire. *Math-École*(numéro 174, octobre 1996), 4-16.

Paul, R., & Elder, L. (2001). *Critical Thinking. Tools for Taking Charge of you Learning and your Life*. Upper Saddle River, NJ: Prentice Hall.

Perrenoud, P. (2002). D'une métaphore à l'autre: transférer ou mobiliser ses connaissances? In J. Dolz & E. Ollagnier (Eds.), *L'énigme de la compétence en éducation* (pp. 45-60). Bruxelles: De Boeck.

Piaget, J., & Inhelder, B. (1974). *La genèse de l'idée de hasard chez l'enfant, Deuxième édition*. Paris: Presses universitaires de France.

Proimos, J., DuRant, R.H., Dwyer Pierce, J., & Goodman, E. (1998). Gambling and Other Risk Behaviors Among 8th- to 12th-Grade Students. *Pediatrics, 102*(2).

Rossen, F. (2001). Youth Gambling: a Critical Review of the Public Health Literature (pp. 59): The University of Auckland, New Zealand.

Savard, A. (2008). *Le développement d'une pensée critique envers les jeux de hasard et d'argent par l'enseignement des probabilités à l'école primaire: Vers une prise de décision.* Thèse inédite. Université Laval, Québec.

Savard, A. (2014). Developing probabilistic thinking: What about people's conceptions? . In E. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives.* (Vol. 2, pp. 283-298). Berlin/Heidelberg: Springer.

Savard, A. (2011). Teaching Citizenship Education through the Mathematics course. In B. Sriraman & V. Freiman (Eds.), *Interdisciplinarity for the 21st century: Proceedings of the 3rd International Symposium on Mathematics and its Connections to Arts and Sciences (MACAS)* (pp. 127-137). Charlotte, NC: Information Age.

Shaffer, H.J., & Hall, M.N. (1996). Estimating the Prevalence of Adolescent Gambling Disorders: A Quantitative Synthesis and Guide Toward Standard Gambling Nomenclature. *Journal of Gambling Studies, 12*(2), 193-214.

Shaughnessy, J.M. (1992). Research in Probability and Statistics: Reflections and Directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-495). New York: Macmillan Publishing Company.

Swartz, R.J., & Perkins, D.N. (1990). *Teaching Thinking: Issues and Approaches*. Pacific Groves, CA: Midwest Publications.

ten Dam, G., & Volman, M. (2004). Critical Thinking as a Citizenship Competence: Teaching Strategies. *Learning and Instruction, 14*(4), 359-379.

Tremblay, G.C., Huffman, L., & Drabman, R.S. (1998). The Effects of Modeling and Experience on Young Children's Persistence at a Gambling Game. *Journal of Gambling Studies, 14*(2), 193-214.

Watson, J.M., & Kelly, B.A. (2004). Expectation versus Variation: Student's Decision Making in a Chance Environment. *Canadian Journal of Science, Mathematics and Technology Education, 4*(3), 371-396.

Wynne, H.J., Smith, G.J., & Jacobs, D.F. (1996). Adolescent Gambling and Problem Gambling in Alberta (pp. 117). Edmonton: Alberta alcohol and drug abuse.

Yinger, R.J. (1980). Can we really teach them to think? In R. E. Young (Ed.), *Fostering critical thinking*. San Francisco: Jossey-Bass Inc.

Students' Language Repertoires for Prediction

David Wagner
University of New Brunswick, Canada
Joseph Dicks
University of New Brunswick, Canada
Paula Kristmanson
University of New Brunswick, Canada

Abstract: Communication about prediction is complex in a number of ways. First, language is by nature recursive — language is an indicator of meaning as well as a force that shapes meaning. Second, the same language used to communicate prediction in uncertain environments is used for other purposes. In this article, we describe how the recursive nature of language impacted the choices we made in a cross-sectional longitudinal study aimed at gaining insight into children's language repertoires relating to conjecture. We then explore some Grade 6 students' communication about prediction to develop insight into their meaning and meaning-making with prediction language. From this we raise questions about interpreting data from such contexts. Finally, we discuss implications for educators.

Keywords: mathematics education, language, degrees of certainty, conjecture, authority.

The understanding of possibility, risk, and certainty, like the understanding of any mathematical idea, is mediated by language. Certain language repertoires are necessary to convey the ideas. At the same time, the language used to describe these ideas shapes the way people conceptualize them. This recursive nature of language compelled us to develop a research project to investigate children's language repertoires in relation to conjecture. Having noted similarities in the language of conjecture and of prediction, we structured the classroom activities and interviews in the project to prompt students to make predictions. In this paper, we focus on our research choices in relation to this endeavour. First, we describe choices we made to gain insight into children's language repertoires. Second, we use some of the data from the project to identify issues relating to interpreting data in the characteristically mathematical contexts of conjecture and prediction.

Moving beyond our academic interest in mathematics education, we will argue that the issues we identify may be significant for understanding everyday experience. In particular, we will raise questions about the impact of mathematics class experiences that involve uncertainty on experience outside the classroom. We will also raise questions about the impact of intertextuality between uniquely mathematical ways of communicating about conjecture and everyday ways of interacting about authority.

The investigation of conjectures (hypotheses) is one of the most important mathematical processes. Much mathematics teaching focuses on enabling students to perform particular mathematical procedures, such as adding fractions, factoring polynomials, and calculating probability. These skills appear as standards in curriculum documents and frameworks (e.g., CCSSO, 2010) that are used by curriculum planners and teachers. Research and professional literature, including curricula (e.g. New

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 246-261

Brunswick Department of Education, 2010) and curriculum frameworks, point to the necessity of students learning these intended outcomes through the exploration of mathematical problems.

When people explore a mathematical problem together, as with mathematical investigations in classrooms, it is necessary to have a way of suggesting an idea before knowing it is true. Rowland (2000) noted the centrality of such conjecture to mathematics, and coined this "space between what we believe and what we are willing to assert" (p. 142) as the Zone of Conjectural Neutrality (ZCN). Because of the recursive relationship between language and experience, the language resources available affect the possibilities for making conjectures.

As our research exemplifies, the language of conjecture shares language that describes probability. Rowland's work refers often to the necessity of expressing uncertainty for conjecture, and he draws heavily on linguistics literature that describes the way people express uncertainty. Our research illustrates the complex relationship between probability, itself an important mathematical concept, and conjecture, which is at the heart of teaching for understanding.

Our interest in language is not aimed to identify correct language. Rather it focuses on the language students use and asks what their language choices might tell us about the way they think about uncertainty. There is a range of English words that relate to uncertainty. Mathematics educators are likely to have particular ideas of what the words mean, which would differ from ideas of others. For example, in addition to everyday use of the word 'risk,' the concept has been studied in the fields of mathematics, psychology, business, and engineering. We find a general consensus that it references probability and uncertainty, especially as they relate to (perceived) consequences (e.g., Slovic, 2000). Our focus in this article is on the meaning and meaning-making we observe while students are confronted with uncertainty.

Communicating About Uncertainty

Our theoretical perspective for this research draws on the work of Vygotsky (e.g., 1962, 1978) and Wertsch (1991) related to the connections between thought and language, and, in particular, the central role that language as social interaction plays in the process of learning.

Nevertheless, we have found it a challenge to avoid deficit framing because of the shaping force of one's language repertoire. Indeed, we suggest that it is not possible to completely avoid deficit framing when analyzing language use. Deficit framing suggests that one's own way of speaking or thinking is superior by evaluating whether or not others have acquired the same skills. In the study of linguistic variation for numbers, which is the only area of mathematics register variation that has been documented significantly, Swetz (2009) pointed out how cultures have been rated on the scope of their number systems. In our research we are more interested in the potential for linguistic variation to open up opportunities to understand mathematics differently. For example, in the context of language repertoires for number, numbers are verbs in Mi'kmaq (Lunney Borden, 2010). Our conversations among ethnomathematicians suggest that this is not uncommon, though in English, numbers are adjectives or nouns. A question warranting attention is how this distinction affects one's conception of counting and arithmetic operations.

We want to bring the same kind of question to the study of different language repertoires for expressing uncertainty. However, at this point, we focus on the range of language strategies within English and French. In this article, we focus on English only. Thus our research addresses the larger question: How does linguistic variation express itself in relation to understanding probability? Because linguistic variation in mathematics (besides the area of number) has not been researched significantly, the discussion requires careful research to move forward.

Reports on research of probability learning have included many examples of language spoken and written by children and their teachers. The same can be said about most subfields of inquiry in mathematics education, in which reporting generally focuses on understanding without significant consideration of the language that mediates this understanding. Morgan et al. (2014) have said that this approach is naïve: "Naïve conceptions of language as a transparent means of transmission of ideas from speaker to listener have been seriously challenged by current thinking about communication" (p. 847). A further naiveté identified by Morgan et al. (2014) is evident in research that views "language as a barrier to learning that must be overcome" (p. 846).

Our review of the subfield of inquiry into probability learning identifies the first kind of naiveté (language as a transparent window) and thankfully not the second (language as a barrier). We position our article as a small first step away from the naïve use of language as a transparent window into probability understanding. It is a small step because we merely problematize apparent meaning-making. The concern we have in our reporting is that it may be taken as a caution to speak and write more clearly, which would relate to the second kind of naiveté. Rather, we will claim that ambiguity is inevitable. However, it is important to understand the nature of this ambiguity.

We see some development of attention to language and related developments in its conceptualization within the subfield of research on probability teaching and learning. Ben-Zvi and associates moved from a naïve conception of language as transparent medium to language in interaction with activity. Ben-Zvi and Arcavi (2001) conducted a study that is similar to ours in many ways. They explored the conceptions of data among children of a similar age to children in our study. In their analysis, they attended carefully to the linguistic distinctions made by their participants. They identified the distinctions among the things said by their participants as representations of different conceptualizations: "The verbal abilities of these students allowed us to follow, at a very fine level of detail, the ways in which they begin to make sense of data, data representations, and the 'culture' of data handling and analysis" (p. 35). This is an example of research that conceptualizes language as a transparent window into understanding. Later, Ben-Zvi et al. (2012) presented a more complex conceptualization of language, as they trace its development in individuals as they interact with increasingly complex probability tasks: "prediction tasks, helped in promoting the students' probabilistic language" (p. 913).

The research we describe in this article is inspired by an unexpected result from an earlier collaboration that was focused on authority in mathematics classrooms (e.g., Herbel-Eisenmann, Kristmanson & Wagner, 2011; Herbel-Eisenmann & Wagner, 2010), for which we drew on Rowland (2000) for interpretation of students' use of modal verbs. The ambiguity of meaning among modal verbs highlighted our attention to the two mathematical phenomena of probability and reasoning. Our current attention to modal verbs was bolstered by the strong attention to modal verbs in literature on additional language teaching and on teaching in multilingual contexts. Thus our interest in uncertainty focuses us on participants' repertoires for expressing modality, especially on the use of modal verbs.

Modality

Modality refers to linguistic tools for expressing degrees of certainty, for example the use of modal verbs like *must* and *could*. "It must be six" is stronger, and thus has higher modality than "It could be six." Rowland (2000) identified assertions that appear without expressions of modality as *root modality*. These assertions, which may be called bald assertions, are stronger than even the highest modal expressions because saying ‘it is six’ does not even recognize the question of other possibilities. Figure 1 illustrates a range of meanings of modal verbs (from Herbel-Eisenmann, Kristmanson & Wagner, 2011, p. 2).

<table>
<tr><td>it is six</td><td colspan="2">high polarity (root modality)</td></tr>
<tr><td>it must be six</td><td rowspan="4">modulated polarity</td><td>high modality</td></tr>
<tr><td>it might be six</td><td rowspan="2">low modality</td></tr>
<tr><td>it could be six</td></tr>
<tr><td>it cannot be six</td><td>high modality</td></tr>
<tr><td>it is not six</td><td colspan="2">low polarity (root modality)</td></tr>
</table>

Figure 1. Range of meanings of modal verbs

Some modal verbs—e.g., ‘can’—are ambiguous. “You can be excused from the table” indicates a degree of obligation; “You can finish the race” indicates ability; “I can help you” indicates inclination; and “It can be a six (because one of the remaining cards in the deck is a six)” indicates probability. When students (or others) hear the word *can*, we wonder what it means to them? Linguists, Martin and Rose (2005), codified these forms of meaning of ambiguous modality language, including categories that describe degrees of certainty relating to usuality, probability, obligation, inclination, and ability (p. 50). Their distinction between usuality and probability is not very clear to us. Rowland (2000) distinguished among the various meanings using the terms *alethic* (or logical), *deontic*, and *epistemic*. Alethic modality is most relevant to communication about probability. For example, in such a context, “the next card *could be* a six” is a statement of certainty if one knows that there is a six in the pile from which cards are being drawn. However, *could* often expresses doubt. Deontic modality references obligation or authority. Epistemic modality references levels of belief and thus is most important in conjecture.

Modality can be expressed using other language strategies, in addition to modal verbs. For example, the adverb ‘probably’ expresses strong confidence short of certainty, and the adverb ‘possibly’ represents the opposite end of the scale, with approximately the same modality (the same level of certainty), but strong confidence that an event is nearly impossible. However, there are no standardized quantifications of these expressions in everyday use. Nevertheless, in quantitative research methodologies certain expressions of modality are taken to have numeric thresholds relating to correlation coefficients, for example. Similarly, in textbooks and other resources for learners of additional languages, modal verbs have appeared with numeric ranges given as percents.

In addition to the alethic and epistemic modality, which relate most closely to probability and reasoning respectively, our analysis will identify modal expressions as referencing emotions — for example, desires and fears. Pratt et al. (2012) demonstrated the influence of emotion in their study of the priority heuristic. The participants’ worries influenced their assessment of probabilistic situations due to empathy for the subject of the context story, which was related to the risk in a medical procedure. One of the conclusions the researchers draw from their study is that “Teaching about risk carries with it certain obligations. We see one pedagogic challenge as sensitising people to their own decision-making, including their emotionally-charged heuristic thinking” (p. 940).

Our analysis will focus on the range of meaning in participants’ language choices, not on the conceptualization of probability. Nevertheless, our data could provide examples of various heuristics and misconceptions identified in the stochastic reasoning literature. Most prominently, we see our participants straddling the boundary between what Fischbein (1975) referred to as ‘primary intuitions’ and 'secondary intuitions'. The primary intuitions are "cognitive acquisitions which are derived from the experience of the individual, without the need for any systematic instruction", and the 'secondary intuitions' are "acquisitions that have all the characteristics of intuitions but . . . are formed by scientific education, mainly in school" (p. 117). The finer-grained distinction developed by Jones et al. (1999) — distinguishing among prestructural, unistructural, multistructural, and relational thinking — could be used to identify student thinking on Fischbein's fuzzy boundary, but we are more interested in distinctions in language than on rating or evaluating children's understanding.

As noted above, we are especially interested in the way children use language to express modality in mathematics contexts (and beyond) because modality is important in conjecture, as noted by Rowland, and to describe uncertainty, and it is also important to understand other points of view, as noted by Shaffer (2006). In our research, we did not aim to look for holes in children's language repertoires. Rather, we focused on attending to the ways they talked about their understanding, to help us see a range of ways to talk about and understand conjecture and uncertainty.

Methodological Choices

The data for our cross-sectional longitudinal study comprise audio- and video-recordings from English-medium and French Immersion instructional contexts in an Anglophone region in Canada. Students worked in groups in class and were subsequently interviewed, extending the group work. At the end of the interviews we asked the students about the meaning of the words they used to describe degrees of certainty: How do the participants distinguish between obligation and probability, as noted above?

For each mathematical context we tried to avoid using specialized mathematical language ourselves. We know from second language acquisition literature that good language learners are generally good at noticing and, subsequently, using the language used in interactions with more able speakers (Long, 1996). Furthermore, people tend to follow the grammatical patterns of their interlocutors, for example in research interviews (Wagner, 2003). We wanted to hear what language skills the children in our research used to communicate their ideas without setting them up with the specialist language to build on. As we struggled to construct problems without use of specialist language, we found that larger narrative contexts made this possible. Other strategies we considered became grammatically awkward.

Our narratives also made the problems accessible to very young children, perhaps partially because of the lack of specialty language, but mostly, we think, because they connect to children's experience. If we were interested in assessing the level of probability knowledge among our participants, we would have to be wary of the effect our narrative contexts could have on supporting their developing conceptualizations. Pratt & Noss (2002) engaged children in micro-world exploration and noted how their understanding developed through their experience within the research. The same caution would apply if we were interested in rating or assessing our participants' language repertoires. As noted by Ben-Zvi et al. (2012), whom we cited above, their participants seemed to develop language for describing probability when engaging in well-constructed tasks involving increasingly challenging questions about uncertainty. However, we are not as interested in rating our participants' understanding and language, as we are in documenting the range of language repertoires used to interact with uncertain contexts.

In addition to embedding our questions in a narrative context, we attempted to avoid specialized uncertainty language when we interviewed participants about their predictions in contexts based on uncertainty. At first, our team agreed it would be acceptable to use a language strategy only after the participant did, not before. This proved extremely difficult; indeed, in the interviews we often used words we intended to avoid, and sometimes used incorrect or awkward structures in attempts to avoid this. After completing most of the first year's classroom and interview interactions, we agreed amongst our team that we should be less paranoid about avoiding specialty language, but thought that this issue might impact the interpretation of the data.

The first year's participants were in Grades 3, 6, and 9. We had them play a modified version of *skunk*, a game often used in the teaching of probability (e.g., Brutlag, 1994; Neller & Presser, 2004). We had them play in pairs so that they would be more likely to talk with each other about their ideas and strategies. We introduced the game with a narrative like this, varying slightly between contexts because we did not script the narrative:

> One day, I was picking strawberries in the forest. After a while, when my basket was quite full, a skunk wandered into the berry patch. I ran away so the skunk would not spray me. I lost the berries in my basket when I ran off.

This narrative also gave a reason for calling the game *skunk*. Participants had a pile of beans (representing the berry patch), a cup (the basket), and a bowl (home). When the researcher rolled the die and called out the number, participants put that number of berries in their basket. A six represented the skunk. When it was rolled, everyone would lose the berries in their baskets. On the other hand, if they "went home" (dumping their beans into their bowl) before the appearance of the skunk, their berries were safe. We played seven rounds — one berry-picking expedition for each day of the week.

We played the game with participants in their classrooms first. The following day we interviewed groups of students and played again but with six cards bearing the numbers one to six instead of the die. The interviewer would not replace the cards into the deck until the deck was completely played out, at which time it would be reshuffled. Thus the participants experienced the difference between independent and mutually exclusive events in probabilistic situations.

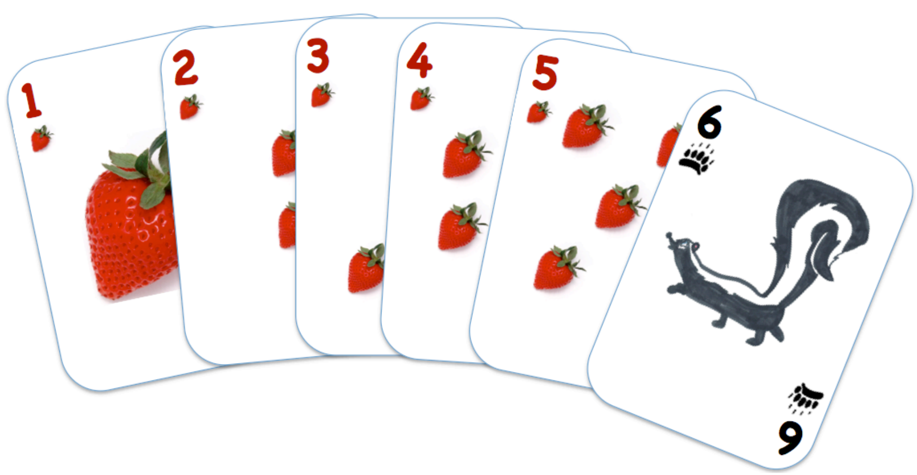


Figure 2: Skunk cards

During the game, the interviewer asked the participants to say why they made their choices about when to "go home." After the game, the interviewer asked participants about specific things they said, asking for clarification on meaning. The camera operator was helpful in this regard, acting as a second interviewer. She or he could make notes on what participants said, which was relatively difficult for the primary interviewer who was busy with the cards and interaction.

For this article, we focus on one interview with four Grade 6 students playing the game of *skunk*. This group of students was not identified by their teacher as exceptional in any way. The school is in an area that has relatively low socio-economic indicators. As noted above, these four students played *skunk* in class the day before, and subsequently one of our research team interviewed them — first playing *skunk* with cards instead of a die, and then asking them about some language meanings. We asked them to play *skunk* in pairs, and they somehow came to an implied agreement that the pairs were competing against each other.

Though we focus on the interview with the four students described above, we make some references in our discussion to other data within the project to illuminate certain findings through comparison. For this reason we describe the second year research prompts as well. Instead of playing *skunk*, participants from Grades 4, 7, and 10 (catching some of the same students as the previous year, one grade earlier) predicted the 50th car on different trains based on the first seven cars. The narrative context of this situation had the researcher tell a story about waiting with a friend for a train at a level crossing, and deciding to predict what kind of car the fiftieth car would be. Trains were then shown using presentation software, with an engine and the first six or seven cars, each labelled with their number, as shown in Figure 3. After students made their predictions about the 50th car, we had the train accelerate and then decelerate to settle on the 50th car. As with the game of *skunk*, we had students work in groups to draw out communication.

Figure 3: First train

The sequences presented to students varied considerably to defy expectations of certain kinds of patterns. The cars were distinguishable by colour and shape — Yellow (Y) cars were rectangular boxcars, green (G) cars were tankers, and blue (B) cars were flatbeds carrying big triangles. Train 1 showed Y,G,B,Y,G,B,Y and continued with a pattern of threes (YGB). Train 2 showed Y,G,Y,Y,G,Y and continued with a pattern that increased the number of Ys before each G — i.e., Y,G,Y,Y,G,Y,Y,Y,G, etc. Of course, the initial seven cars could have suggested a pattern of threes (YGY) similar to the previous train — i.e., Y,G,Y,Y,G,Y,Y,G,Y, etc. For this train, we stopped the train at around the 25th car to let students reconsider their predictions. We invited students to tell us their reasoning whenever possible. Train 3 showed B,G,B,B,G,G, etc. and continued with B, B, B, G, G, G, etc. with increasing groups of B and G. The interviews on the following day started with Train 4 showing Y,B,G,Y,Y,B,G. It continued with groups of four (YBGY) — i.e., Y,B,G,Y,Y,B,G,Y etc. Train 5 started with Y,B,G,B,P,B,Y and continued with a random collection of cars, in which the colours started to misalign with the shapes and new kinds of cars appeared, including ones carrying animals. As with Train 2, we stopped train 5 at around the 25th car so we could talk with the students as they reconsidered their predictions. In addition to the confounding randomness of the fifth, (perhaps "avant-garde") train, there was no 50th car — it had only 42 cars. As with *skunk*, we ended these interviews with questions about distinctions among various language choices we heard the students use.

Language Used to Express Uncertainty

The four 11- and 12-year olds in the group whose interview we focus on for this article show considerable language repertoires, which we found to be the case for even the most mathematically and linguistically novice students in this project, even the Grade 3 French Immersion students who were in their first year of French-medium learning. As noted above, we were the most careful about and attentive to modal verbs in our analysis because of our earlier research and teaching work, but we identify other ways to describe levels of certainty as well. We present here a narrative of the game played in the interview, followed with more detailed analysis of the discussion about the meaning of some key modal verbs.

Playing skunk

We began the interview with a question about the dice version of the skunk game played in class the day before: "You know how the skunk came on the six, what if the rule was that the skunk came on the three instead of the six? Would you get more berries or less?" (turn 30). The students demonstrated either a misconception (possibly due to the prevalence of sixes and the paucity of threes on the day before) or based their answers on primary intuition. Nevertheless, they gave a few expressions that they seemed to feel had similar meaning. Chris discounted the "difference" between the two sets of rules by saying "It wouldn't really make much of a difference" (turn 31). Dale needed no special language to agree, simply saying, "No" (turn 32). Chris added the expression "an even chance" with an indicator of reasoning — 'because': "Because they are all an even chance" (turn 33). Finally, Terry said, "it's just the roll of the dice" (turn 38), apparently using the word 'just' to suggest that nothing different happens with either set of rules. Thus there seemed to be four different ways of expressing the idea that the probability remained unchanged.

We then moved to playing *skunk* with the cards. After shuffling the deck of six cards, the first cards played were the 2, the 1, and then the skunk. Each pair of students lost the three berries collected before the skunk had arrived. The three cards (2, 1, and skunk) were laying face up on the table. The interviewer, holding the remaining three cards, gestured that the students should decide to stay or go home. Dale pointed at the skunk card and said, "Well, the skunk is right there" (turn 74). Dale needed no special language to indicate certainty — the gesture along with the bald assertion sufficed.

The interviewer then drew attention to the next unrevealed card: "Okay, so what about this one?" (turn 84). Chris responded with, "That would be like around five." This expression employs a mix of uncertainty language. "Would be" expresses prediction. "Like" lowers the modality even more, as does "around" (though it is more normally used to describe estimation).

The 4, and the 3 were played next so five cards now laid face up on the table (2, 1, skunk, 4, 3). Chris asked, "Do you have a second skunk in there?" (turn 95). The interviewer was surprised by this question and responded, "I showed you the cards before. What cards were they?" (turn 96). Chris again expressed doubt, "Trickster" (turn 97). Chris seemed to think the interviewer might have secretly switched one of the cards.

The interviewer played out the deck, picked up the cards and shuffled them. This time the skunk came out first. One of the pairs of students went home before the skunk was played and thus averted losing their berries. The interviewer asked, "Did you know? Did you know that this was the skunk?" (turn 124). Chris said, "No, I just, kind of had a feeling" (turn 126). Dale said, "had a feeling" (turn 127) simultaneously with Chris. This reference to having a feeling did not seem to be associated with emotion. Rather, it seemed to be a strategy for expressing the act of prediction. Interestingly, both Chris and Dale introduced the expression simultaneously.

By contrast, close to the end of the game, emotion became palpable and the students talked about risk. Chris said, “Don’t want to take our chances” (turn 160). Leslie then reacted with emotion to the emergence of the skunk as the first card of the newly shuffled deck (this was the second time the skunk was the first card in the deck): “Oh, come on!” (turn 164) laughing with what appeared to be a mix of delight and frustration. After the five remaining cards were played out, for which the students wisely elected to stay in the berry patch, the cards were reshuffled and the students discussed their strategy in earnest. Leslie said, “I’m scared” (p. 191), expressing emotion but perhaps also expressing awareness of the possibility of the skunk appearing early in the deck again. Terry assessed the risk of loss in relation to the possibility of catching up to the other pair: “I mean, if you guys still stay, then we really don’t have anything to lose” (turn 198). Later Terry added, “One of us won’t lose everything” (turn 204), and Leslie responded, “It is probably going to be us” (line 205). This instance of ‘probably’ seemed to be based both on calculation and fear.

As noted in the above narrative, for each idea there seemed to be more than one way of expressing it, including very simple statements that indicated agreement with another utterance. Also noted above, the prediction, conjecture, and emotion seem to be intertwined in the students’ utterances.

Discussing meanings of modal verbs

We turn now to analysis of the modal verbs. The modal verb *have* expresses high modality because it refers to events that must occur. The interviewer used it first (though trying to avoid doing so) in turn 111, but it did not get used again until turn 229 when Chris talked about the difference between playing skunk with cards and with the die: “It’s easier this way because when the skunk first came you just don’t have to worry.” It wasn’t used again until the interviewer asked questions about its meaning. Here is an abbreviated version of that discussion (omitting diversions).

> 319 *Interviewer*: [Yesterday] I heard Terry say when you're working in your groups, “Do we both have to write this down?” So what’s the difference between “it has to be the skunk” and “she has to write this down”? Is the “has to” the same? “This has to be the skunk.” “She has to write it down.” Do you notice a difference between them?
>
> ...
>
> 346 *Terry:* Do we both need to, like, do we both need to write it down?
>
> 347 *Interviewer:* No, but it’s a proper use of the word. But is it the same as “this has to be the skunk”?
>
> 348 *Terry:* No.
>
> 349 Interviewer: No? Why not?
>
> 350 *Terry:* Because you know it has to be.
>
> 351 *Dale:* It absolutely has to be.
>
> 352 *Interviewer:* It absolutely has to be.
>
> 353 *Terry:* Yeah.
>
> 354 *Interviewer:* But when asking “do you have to” it’s not absolutely.
>
> 355 *Terry:* No, yeah.
>
> 356 *Interviewer:* Okay

357 *Dale:* Because the fire bell or something could ring or something and you all go outside and you don't have to write it down.

358 *Interviewer:* Don't have to write it down but if the fire bell rung this would still be the skunk.

359 *Dale:* It would still be the skunk.

We note that, to clarify meaning, the students introduce new vocabulary that was not part of the interview up to this point. Terry used the modal verb structure "need to" to emphasize the necessity of "have to." Dale introduced the adverb *absolutely* to further emphasize this sense. The students distinguished between instances of 'have to' depending on context.

We had a similar conversation about the modal verb *can* which had been used in its various forms, including *can't*, by the students in the interview. We started this part of the conversation by referencing Dale's writing in class earlier. When asked what is the greatest number of berries they could get in a day, Dale had written, "You can get any number because it could just keep going." (This was with playing skunk with a die.) The researcher also referred to Dale saying in the interview that it is different with the cards because "we can't keep going." Again, this is an abbreviated version of the conversation that ensued.

391 *Interviewer:* That *can't* – If you're wanting to go visit your friend, and your mother or father says that you can't go over to your friend's house, is it the same kind of *can't*

392 *Terry:* No, that means you're not allowed.

393 *Interviewer:* Not allowed. So how do you know, if you teacher says that you can't do something, whether she is teaching your something?

394 *Terry:* It means no, you're not allowed to.

395 *Interviewer:* You're not allowed to.

396: *Terry:* Yeah.

397 *Interviewer:* Of how do you know it's not the kind of *can't* that Dale said? Where it just can't possibly happen? How can you tell the difference?

398 *Terry:* By the way she says it.

399 *Chris:* Yeah.

...

418 *Camera operator:* When you said earlier "you can't win," which one is that closest to? Remember, when you looked at your basket and you said, "Oh, we can't win." Is that like the "you're not allowed" or is it.

419 *Terry:* It would be you can't.

420 *Leslie:* You don't.

421 *Terry:* Like you, it's impossible, like.

422 *Leslie:* Yeah, it's impossible.

423 *Terry:* Well, it was because if you added it all up, the skunk...

424 *Dale:* You'd only get, like, fifteen.

425 *Terry:* The skunk would have come.

426 *Chris:* Yeah, you'd only get fifteen so if the skunk is that

427 *Leslie:* Or seventeen.

428 *Dale:* No, because we had thirty-five and then I counted them all up form the five and we still wouldn't have enough.

429 *Interviewer:* Okay.

430 *Terry:* Because the skunk was gone.

431 *Interviewer:* It would have been impossible.

432 *Terry:* Yeah, yeah.

433 *Interviewer:* So if someone says can't, … if I told you that you can't divide by zero in a lesson on dividing would you think that that means that you're not allowed to or that it is impossible to do?

434 *Chris:* That it is impossible.

435 *Interviewer:* Why would you think that?

436 *Terry:* Because you can't divide by zero.

437 *Interviewer:* Why can't you?

438 *Terry:* Because it is impossible.

439 *Interviewer:* How do you know?

440 *Chris:* Because you can't.

441 *Terry:* Because you can't.

442 *Chris:* If it is zero, you can't put it in any groups.

In this case, Terry introduced the adjective *impossible*, to clarify the meaning of *can't*. No one had used the word before this in the interview. As with "have to", the students distinguished among instances of *can* and *can't* based on context. During and after this interview, we wondered how students could make this distinction for instances in which they do not know a convincing logical argument for the assertion. With the example of division by zero, the students now knew that it is impossible, but how might they have thought about it the first time they heard their teacher say, "you can't divide by zero"?

Overview of linguistic strategies

The students introduced three adverbs/adjectives that indicate degrees of probability into the interview. The word *probably* was first used by Leslie and not used again by others. When Leslie and Terry were considering whether or not to make the same choice about going home or not as the other group, Terry remarked, "One of [our groups] won't lose everything and the other would" (turn 204), and Leslie replied, "It is probably going to be us" (turn 205). The adverbs *absolutely* and *impossible* (sometimes an adjective) came up in the conversations about language choices when the students were trying to explain what the modal expressions meant, as noted above.

Other modal verbs used included *would*, which was first used (accidently) by the interviewer and used liberally later by the students, and *may* as in "you may be able to win" (Dale, turn 263). Another specialized linguistic form used by a student was the if-then statement, first used by Chris: "If it was two numbers, then it would make a difference" (turn 39). This was in the context of discussion about the playing *skunk* with a die.

In addition to the relatively specialist terminology for modality (the modal verbs and adverbs), students expressed degrees of certainty in other ways. Terry introduced the modal expression "I think" in a conversation about playing skunk with the die. The researcher had asked if the number of berries they gathered would be different if the skunk came on a one instead of a six, to which Terry replied, "I think it would because we roll the one a lot" (turn 45). Terry introduced another expression to describe the differences between playing skunk with cards and with the die. In turn 236 Terry said, "You never know what is going to happen (with the die)." Terry also said, "the odds are harder" (line 273) when the probability of success became lower. Dale was inventive too, and used the expression "I had a feeling" (turn 126) after "going home" to stay safe. This statement was in reply to the researcher asking, "Did you know that this was the skunk?"

Finally, the absence of any modal expressions is significant in the consideration of modality as well. The use of bald assertions can replace strong modal verbs or adverbs. Dale said, "the skunk is right there" (line 74) while pointing at the skunk card, as yet unrevealed but evidently the skunk by deduction. We might expect "the next one has to be the skunk" or "I am certain that the next one is the skunk" but the bald assertions serves the same function. Chris did the same on line 82 saying "it's there." In this interview (and others), there were many instances of this method for expressing certainty.

Discussion

The four students in the interview described above demonstrated a wide repertoire of language for expressing degrees of certainty. Each of them used a range of expressions, and each of them introduced expressions that no one else had used before. Terry was the most talkative in the discussions about language meaning, but we caution that it would be unwarranted to make conclusions in comparison to the others on this basis. Many of the expressions introduced by the students came late in the interview, which tells us that if the interview had been shorter, we would not have known whether or not the students had these expressions in their repertoires. This serves as an exemplary caution against deficit-based assessments. The development of increasing linguistic strategies may also raise questions about the work of Ben-Zvi et al. (2012), cited above — were the participants in that study increasing their linguistic strategies or simply waiting to employ their strategies when they are needed? Both are reasonable explanations.

Another phenomenon that challenges deficit-assessment is that when one student said something, there was no need for the others to say it again or even speak about it unless they disagreed. We cannot assume someone does not possess certain language simply because they do not use it. However, we can claim that a student has an expression in their language repertoire if they introduce it. This is why we went to the lengths that we did for structuring our prompts carefully.

Related to this, we note that if students use an expression that has just been used by the teacher or interviewer, perhaps in a recent class, or in the interview itself, the student use may not be fully independent. They may not be able to use the expression autonomously at a later time or in a different context.

In addition to using (and introducing) specialist language, the students in the interview at times demonstrated ability to convey their meaning using very limited technical language. In particular, they could make their ideas clear when talking about the extremes of certainty — when events were impossible or certain. The more specialised language seemed to be relied upon either for describing events that were somewhere between impossible and certain (aligning with a result from Ben-Zvi et al. (2012)), and for clarifying meaning on the extremes when pressed to do so.

As noted above, Rowland (2000) introduced the idea of the zone of conjectural neutrality to describe language that specifies degrees of certainty, which is "in defiance of the cultural norm that the pupil is judged to be 'right' or 'wrong'" (p. 211). He claimed it to be helpful for a conjecturing atmosphere. We note that the same terminology is used to describe probability, and thus specialized modality language can defy situations in which predicted results may be between impossibility and certainty. We have only begun to consider the implications for pedagogy considering the phenomenon that language is shared for both conjecture and probability spaces.

This brings us to discussion of the second research context, which was set up to be similar to but distinct from the game of skunk — a twist on the context. In both contexts, students were making predictions. What is the difference between a train and a pile of cards, both of which are sequences of physical objects? One difference is that the cards are shuffled in front of the students and train cars are sequenced with some sort of intention in advance. Nevertheless, our experience of real trains is that the sequence of cars seems to be quite random, or in groups (e.g., the boxcars first, followed by a bunch of tankers, followed by a few flatbeds, and finally the rest of the tankers). We have never seen trains with patterns similar to the ones introduced in our research — patterns like yellow boxcar, green tanker, blue flatbed, yellow boxcar, green tanker, blue flatbed, etc. A Grade 4 student in the second year of research involving the trains became increasingly frustrated with the rest of the class identifying what the 50th car would be. This student kept saying that it is impossible to know, while the class continued to ignore him. This student refused to make predictions. This reminds us of an observation noted by both Falk (1981) and Chernoff (2009) — that people often see randomness when it is not present, and see order when randomness is present.

This tension also points to the presence of some sort of pedagogical contract in which students generally expect intention from their teachers. We would suggest that this contract extends to researchers, whom, from our experience, are associated with teachers. Even in the game of *skunk*, when the interviewers showed all the cards to the students and shuffled the cards directly in front of them, the students sometimes expected some kind of lesson — the appearance of a second skunk card, for instance (for example, Chris expressed this fear in the example above). With the trains the phenomenon was more obvious; the students (with some exceptions, most notably the Grade 4 student noted above) assumed that the patterns would continue even though the researcher and teacher never said that these were patterns and the described context was one of a real-life train. The apparent frustration displayed by almost all the participants when they saw the fifth train (the random, avant-garde train) made clear to us the students' expectations for pattern. There is something about the transposition of a narrative into a mathematics classroom that changes it to a scenario in which everything should be predictable (and known by the teacher, or researcher). We suggest that this transposition may confound some claims in the literature based on classroom interactions or research interviews about probability.

In our research project, student predictions were based on both the probabilities inherent in the given scenarios and the students' second-guessing of teacher/research choices in constructing scenarios for pedagogic or other reasons. This raises questions about how students experience probability learning. Uncertainty in the mathematics classroom is experienced differently than it is outside the classroom. Furthermore, we note that the language of conjecture shares language with probability and risk-related emotion, and so we wonder whether this ought to confound similarly our understanding of the way students experience proof and reasoning.

Finally, we turn our attention to implications beyond the classroom. Increasingly significant social phenomena, such as climate change, involve both calculations of risk, which are based on assumptions, and conjectures (hypotheses). The fact that risk calculation and conjecture share terminology may complicate communication about such social phenomena. Furthermore, language used to express risk calculation and conjectural language of certainty are also used to express authority, as demonstrated in the above conversation about authority — notably the discussion about the modal verbs ‘have to’ and ‘can’t’. When people in the public sphere who appear to be scientific make claims that sound authoritative, how are listeners to know whether these claims are warranted expressions of calculation-based certainty? It is incumbent upon mathematics teachers to be aware of these shades of meaning and the risk of ambiguity on such important social issues.

This brings us to the question about what educators might do in the face of the ambiguity in this language. Pratt et al. (2012) conclude their study of emotion-laced contexts of risk assessment saying, “Teaching about risk carries with it certain obligations. We see one pedagogic challenge as sensitising people to their own decision making, including their emotionally-charged heuristic thinking” (p. 940). We suggest that this imperative is warranted not only for discussion of emotion but also for discussion of linguistic ambiguities. As educators we are obligated to help students and other educators become aware of the meaning associated with the language of prediction (uncertainty), emotion, authority, and reasoning. The ambiguity of this meaning is probably inevitable however. First, because uncertainty is by nature worrisome and thus triggers emotion. Second, people who want to establish authority will co-opt the language of logic to emphasize what they consider to be necessity. Third, reasoning requires the acknowledgment that one might be unsure of an idea. Developing a fuller repertoire of language to express ideas may help people negotiate meaning, but will probably not make meaning and meaning-making entirely clear because of the inherent connections among these concepts.

Perhaps an appropriate way to close this article is to employ one of the ambiguous expressions discussed here — good educators *have to* (or must) make their students aware of the ambiguity in prediction and reasoning language. We have to because it is a moral obligation. And we have to because there is no way around this awareness when we aim for clarity.

Acknowledgements

This is an elaboration of a paper presented at the 9th Congress of European Research in Mathematics Education (2015). The research was supported by Canada’s Social Sciences and Humanities Research Council (Grant title: “Students’ language repertoires for investigating mathematics”). Opinions, findings, and conclusions or recommendations expressed here are the authors’ and do not necessarily reflect the views of the granting bodies.

References

Ben-Zvi, D. & Arcavi, A. (2001). Junior high school students’ construction of gloval views of data and data representations. *Educational Studies in Mathematics,* 45 (1), 35–65.

Ben-Zvi, D., Aridor, K., Makar, K. & Bakker, A. (2012). Students’ emergent articulations of uncertainty while making informal statistical inferences. *ZDM: The International Journal of Mathematics Education*, 44 (7), 913–925.

Bley-Vroman, R. & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. Gass, & A. Cohen (Eds.). *Research methodology in second language acquisition* (pp. 245-261). Hilldale: Lawrence Erlbaum.

Brutlag, D. (1994). Choice and chance in life: the game of SKUNK. *Mathematics Teaching in the Middle School*, 1 (1), 28-33.

CCSSO (2010). *Common core state standards for mathematics*. Wasington D.C.: National Governors' Association Center for Best Practices and Council of Chief State School Officers (CCSSO).

Chernoff, E. (2009). Sample space partitions: an investigative lens. *Journal of Mathematical Behavior*, 28 (1), 19–29.

Falk, R. (1981). The perception of randomness. In *Proceedings of the fifth conference of the International Group for the Psychology of Mathematics Education* (pp. 222–229). Grenoble: University of Grenoble.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Herbel-Eisenmann, B. & Wagner , D. (2010). Appraising lexical bundles in mathematics classroom discourse: Obligation and choice. *Educational Studies in Mathematics*, 75 (1), 43-63.

Herbel-Eisenmann, B., Kristmanson, P., & Wagner, D. (2011). Modality in French immersion mathematics. *ICMI Study 21 Conference: Mathematics Education and Language Diversity*. Sao Paulo, Brazil.

Jones, G., Langrall, C., Thornton, C., & Mogill, A. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, 30 (5), 487-519.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In Ritchie, W; Bhatia, T. (Eds.). *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

Lunney Borden, L. (2010). *Transforming mathematics education for Mi'kmaw students through mawikinutimatimk*. Unpublished Ph.D. dissertation, Fredericton, NB: University of New Brunswick.

Morgan, C., Craig, T., Schuette, M., & Wagner, D. (2014). Language and communication in mathematics education: an overview of research in the field. *ZDM: The International Journal of Mathematics Education*, 46 (6), 844-854.

Neller, T. & Presser, C. (2004). Optimal play of the dice game Pig. *The UMAP Journal*, 25 (1).

New Brunswick Department of Education (2010). *Mathematics grade 6 curriculum*. Fredericton: Department of Education, New Brunswick.

Pratt, D., & Noss, R. (2002). The micro-evolution of mathematical knowledge: the case of randomness. *Journal of the Learning Sciences*, 11(4), 453–488.

Pratt, D., Levinson, R., Kent, P., Yogui, C., & Kapadia, R. (2012). A pedagogic appraisal of the priority heuristic. *ZDM: The International Journal of Mathematics Education*, 44 (7), 927–940.

Rowland, T. (2000). *The pragmatics of mathematics education: Vagueness in mathematical discourse*. London: Falmer.

Shaffer, B. (2006). Deaf children's acquisition of modal terms (pp. 291-313). In B. Schick, M, Marschark, and P. Spencer (Eds.). *Advances in the sign language development of deaf children*. New York, NY: Oxford University Press.

Slovic, P. (2000). *The perception of risk*. London: Earthscan.

Swain, M. (1995) Three functions of output in second language learning. In G. Cook & B. Seidelhofer (Eds.). *Principle and practice in applied linguistics: studies in honor of H.G. Widdowson* (pp. 125-144). Oxford, U.K.: Oxford University Press.

Swetz, F. (2009). Culture and the development of mathematics: An historical perspective. In B. Greer, S. Mukhopadhyay, A. Powell and S. Nelson-Barber, Sharon (Eds.), *Culturally responsive mathematics education* (pp. 11-42). New York: Routledge.

Vygotsky, L. (1962). *Thought and language*. (E. Hanfmann & G. Vakar, Eds. and Trans.). Cambridge, MA: The M.I.T. Press.

Vygotsky, L. (1978). *Mind in society*. Cambridge: Harvard University Press.

Wagner, D. (2003). Teachers and students listening to themselves. In N. Pateman, B. Dougherty & J. Zilliox (Eds.). *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th Conference of PME-NA*, Honolulu, USA, vol. 4, pp. 355-363.

Wertsch, J. (1991). *Voices of the mind*. Cambridge: Harvard University Press.

The Role of Probabilistic Reasoning Abilities on Adolescent Risk Taking

Maria Anna Donati
University of Florence, Italy
Francesca Chiesi
University of Florence, Italy
Caterina Primi
University of Florence, Italy

Abstract: The aim of this work was to investigate the role of the cognitive system and the affective system on adolescents' risk taking in gambling tasks characterized as different on the basis of information given to decision makers. In Study 1, we explored the role of probabilistic reasoning and sensation seeking on decision making in a *non-risky context* (*Non-Gambling Task*) and a *risky context* (*Gambling Task*) in which no preliminary information were given to participants. Results showed that adolescents referred to probabilistic reasoning only in the *Non-Gambling Task*. In Study 2, we explored the role of probabilistic reasoning and sensation seeking in risky situations with preliminary information given to participants. Specifically, we compared a risky context characterized by *high*-emotional arousal (*Game of Dice Task*), in which a feedback after each decision was given, with one characterized by *low*-emotional arousal (*Game of Dice Task* – Modified version), without feedback. Probabilistic reasoning characterized adolescents' decision making regardless of feedback. Findings showed that adolescents' decision making was solely linked to the cognitive system in the non-risky situation, and the affective system overcomes the cognitive system in situation of risk. Moreover, providing information about the task might interfere with the imbalance between the two systems.

Keywords: adolescents, risk taking, gambling, decision making, probabilistic reasoning, sensation seeking.

Introduction

Adolescence is characterized by risky behavioral decisions such as driving without seatbelts, carrying weapons, using illicit drugs and alcohol, and engaging in unprotected sex (Reyna, Chapman, Dougherty, & Confrey, 2012). Recent neurobiological models suggested that two distinct brain systems (i.e., dual-system theory) are involved in decision making and risk taking in adolescence (see Sommerville, Jones, & Casey, 2010, for more details): A *cognitive control system* or cognitive system and a *socio-emotional system* or affective system (Chein, Albert, O'Brien, Uckert, & Steinberg, 2011). Some differences can be identified between them. While the cognitive system is the neural basis of *deliberating processing*, which is effortful and controlled, and operates according to formal rules of logic (Weber, Shafir, & Blais, 2004), *affecting processing* is spontaneous and automatic, and operates by principles of similarity and contiguity. The cognitive system is also the neural basis of *inhibitory control*, a mechanism that can block affective impulses and therefore enables deliberative decision making even in affect-charged situations (Cohen, 2005; Knock & Fehr, 2007; McClure, Laibson, Loewenstein, & Cohen, 2004). On the contrary, the affective system influences behavior by *affective impulses* (Weber et al., 2004). Finally, the two systems have been shown to mature at different speeds. The affective system matures rapidly at puberty, whereas the cognitive system matures later and more gradually over the course of adolescence (Casey, Jones, & Hare, 2008; Steinberg, 2008).

During adolescence, the two systems affects youth's decision making process. In fact, the cognitive system supports decision making process through the inhibition of impulsive behavior, whereas the affective system relies on subjective evaluation that only takes into account potential

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 262-274

rewards of the risky choices. Due to both the relative maturity of brain structures (i.e., amygdala, ventral striatum, nucleus accumbens) that govern the affective system and the immaturity of the cognitive system (i.e., lateral prefrontal cortex), adolescence is a unique time period in which there is an imbalance between these two distinct brain systems (e.g., Casey et al., 2008; Sommerville et al., 2010). Due to this imbalance, adolescents may be more likely to take risks when compared to adults (Chein et al., 2011). In fact, according to these kind of explanations, adolescent risk taking is the result of a competition between the philogenetically younger cognitive system and the phylogenetically older affective system (Casey et al., 2008; Cohen, 2005; Steinberg, 2008).

Based on dual-system models' research, two relevant claims describing decision making in adolescence can be drawn. The first one is that deliberating processing abilities mature earlier than the potential for inhibitory control in the cognitive system (Keating, 2004; Kuhn, 2006). The second one is that the difference in maturational speed between the cognitive and the affective system is assumed to result in a developmental imbalance between the systems during adolescence, with the affective system being easily triggered, for example by the expectation of a reward (Galván et al., 2006; Galván, Hare, Voss, Glover, & Casey, 2007) or the presence of peers (Gardner & Steinberg, 2005). Thus, the very active affective system is not yet sufficiently counterbalanced by the still-maturing cognitive one.

Following these claims, two consequences can be derived, as follows. First, as deliberating processing abilities mature earlier than the potential for inhibitory control in the cognitive system, adolescents can be expected to show a good functioning of deliberating processes in the absence of affective involvement. On the contrary, they can be expected to do not control affective impulses in situations of affective involvement, despite good deliberating processing abilities. Second, since the affective system is not yet sufficiently counterbalanced by the still-maturing cognitive system, adolescents can be assumed to be more susceptible to risk taking in situations of higher emotional arousal.

Starting from these premises, the general aim of this work was to investigate the role of the two systems on adolescents' risk taking in different probability-related decision making gambling tasks (in line with literature: see Figner, Mackinlay, Wilkening & Weber, 2009; Schonberg, Fox, & Poldrack, 2011). In fact, one such example of risky decision making where the imbalance between cognition and affect is evident is in gambling.

Excessive gambling has become a growing problem among adolescents and a recent review found alarming rates of problematic and pathological gambling (see Scholes-Balog, Hemphill, Dowling, & Toumbourou, 2014). Research has suggested that *cognitive* factors based on misunderstanding the nature of probability including a set of erroneous beliefs, irrational thoughts, and misperceptions (e.g., Delfabbro, Lahn, & Grabosky, 2006; Delfabbro, Lambos, King, & Puglies, 2009; Donati, Primi, & Chiesi, 2013; Turner, Zangeneh, & Littman-Sharp, 2006), and *affective* factors, such as sensation seeking (e.g., Donati et al., 2013; Gupta & Derevensky, 1998; Nower, Gupta, & Derevensky, 2004), are associated with excessive gambling. Specifically, adolescent pathological gamblers, relative to non-pathological gamblers, are more prone to mistaken views about randomness and erroneous probabilistic reasoning, tend to hold erroneous beliefs about their chance of winning, are susceptible to biases related to gambling outcomes, and have higher levels of sensation seeking.

Since the cognitive and the affective systems seem to represent respectively 'thinking' and 'affect' domains (Reyna & Rivers, 2008), we considered probabilistic reasoning ability as *'thinking variable'* and sensation seeking as '*affect variable'*. Given that risk taking was measured through gambling tasks, we chose those variables on the basis of the above-described predictive role of probabilistic reasoning (e.g., Delfabbro et al., 2006; Delfabbro et al., 2009; Donati et al., 2013;

Turner et al., 2006) and sensation seeking (Donati et al., 2013; Gupta & Derevensky, 1998; Nower et al., 2004) on adolescent problem gambling behavior.

We also aimed to explore the role of the two systems on adolescent risky decision making in gambling tasks characterized as different on the basis of information given to decision makers. More in detail, from Brand's model of decision making under risk (e.g., Brand, Kalbe, Labudda, Fujiwara, Kessler, & Markowitsch, 2005; Schiebener, Zamarian, Delazer, & Brand, 2011), we know that different kind of decision making situations can occur in everyday life. Specifically, this model claims that as the level of information arises, the risk involved in the decision-making situation becomes more explicit: Decision making under *explicit* risky conditions occurs when decisions can be made on the basis of some knowledge about the situation and associated consequences. On the contrary, when a person cannot know these preliminary information, decision making occurs in a situation of *implicit* risk taking.

Following Brand's model of decision making under risk, one of the characteristics that differentiates risk taking tasks was the presence of explicit preliminary information about task contingencies (including for example the presentation of the number of the task trials and the explanation of potential consequences of each choice option). Based on this claim, we analyzed *implicit* situations of risk, intended as *non-informed* risky situations, since no preliminary information about the task contingencies were given (Study 1), and *explicit* situations of risk, intended as *informed* risky situations, since preliminary information about the task were given (Study 2).

Study 1

The aim of the Study 1 was to explore the role of the cognitive system and the affective system on adolescents' decision making in an implicit situation of risk, i.e. a gambling condition without preliminary information. More specifically, we wanted to explore whether the role of probabilistic reasoning ability on adolescents' decision making changed between a *non-risky context* (non-gambling condition) and a *risky context* (gambling condition). Our hypothesis was that adolescents would used their probabilistic reasoning abilities only in the non-gambling condition, whereas sensation seeking would affect adolescents' decision making in the gambling condition.

Concerning the measurement of decision-making abilities, in order to analyze adolescents' choices in non-risky and risky situations, we developed two versions of an equivalent decision-making task. Specifically, in the first version of the task, adolescents were solely requested to reason in probabilistic terms on a series of coin tosses sequences, while in the following version, they were asked to bet money on the same outcome sequences. Risky decision making was conceptualized as the tendency to commit the *gambler's fallacy*.

The gambler's fallacy is one of the biases related to the representativeness heuristic, which indicates a tendency for people to base their judgment of the probability of a particular event on how much it represents the essential features of the parent population or of its generating process (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). This fallacy occurs when individuals believe that even short strings of random events must correspond with their perception of what constitutes randomness, leading to beliefs that particular outcomes are "due" (Tversky & Kahneman, 1971). For example, when tossing a fair coin, after a series of heads, people have the feeling that a tail should follow, because this corresponds more to their expectation of having a mix of heads and tails, rather than a long sequence of just heads (Morsanyi, Primi, Chiesi, & Handley 2009). We considered the tendency to commit the gambler's fallacy as a risk-taking measure since several studies showed that the susceptibility to commit this bias characterized adolescent problem gamblers (e.g., Delfabbro et al., 2009; Donati et al., 2013; Skoukaskas & Satkeviciute, 2007; Turner, Macdonald, Bartoshuk, & Zangeneh, 2008).

Method

Participants

Participants were 148 adolescents (62% males, mean age=15.9 years, *SD*=1.59) who attended high school in Italy (Tuscany). Written informed assent was provided by students and written informed consent was provided by the parents if the student was a minor.

Instruments

Measures. Probabilistic reasoning ability was measured with the *Probabilistic Reasoning Questionnaire* (PRQ, Primi, Morsanyi, & Chiesi, 2014). Using Item Response Theory (IRT), the PRQ was designed to measure proportional reasoning and basic probabilistic reasoning skills, which are deemed necessary to reason normatively and avoid heuristic strategies. The scale consisted of 16 multiple-choice probabilistic reasoning questions. Items included simple, conditional, and conjunct probabilities, and data were presented both in frequencies and percentages (for examples: "*A ball was drawn from a bag containing 10 red, 30 white, 20 blue, and 15 yellow balls. What is the probability that it is neither red nor blue?*" a. 30/75; b, 10/75; c.45/75; and "*60% of the population in a city are men and 40% are women. 50% of the men and 30% of the women smoke. We select a person from the city at random. What is the probability that this person is a smoker*?" a. 42%, b. 50%, c. 85%). A single composite score was computed based on the sum of correct responses.

Sensation seeking was measured through the *Brief Sensation Seeking Scale* (BSSS, Hoyle, Stephenson, Palmgreen, Lorch, & Donohew, 2002; Italian version: Primi, Narducci, Benedetti, Donati, & Chiesi, 2011). It contains eight Likert-type items using a 5-point scale ranging from *strongly disagree* to *strongly agree*, yielding a maximum score of 40. Higher scores represent high levels of sensation seeking. An example of an item is *"I would love to have new and exciting experiences, even if they are illegal"*. Past studies have shown that the BSSS has adequate reliability and validity (Hoyle et al., 2002; Primi et al., 2011).

Tasks. In order to compare non-gambling and gambling situations, two versions of an equivalent paper and pencil task were developed (Table 1).

	Non-Gambling Task	**Gambling Task**
Coin Toss Sequences	What is the likelihood of Tails at the 7th toss?	You have €10. How much do you be on Tails at the 7th toss?
H-H-T-H-T-T		
T-H-T-H-H-H		
H-T-H-T-T-T		
H-H-T-T-T-T		
T-H-T-T-H-H		
T-T-H-H-H-H		

Table 1. Non-Gambling and Gambling Tasks

The first version was the *Non-gambling Task*. In this task, participants were presented with six different outcome sequences consisting in six coin tosses (T = Tails, H = Heads). For each sequence, they were asked to indicate the likelihood of Tail if a seventh toss would be made. One point was given to each wrong response so that higher scores corresponded to low normative reasoning. After participants had compiled this first task, the *Gambling Task* was administered. In this task, participants were presented with the same outcome sequences as the *Non-gambling Task*, but, for each sequence, they were asked to indicate how much money, from a minimum of €0 to a maximum of €10 (available for each sequence), would bet on Tails if a seventh toss would be made. A net score was calculated by subtracting the average amount of money bet on the first, the third, and the fourth sequence (i.e. those sequences which make more likely to bet on Tails according to the gambler's fallacy bias) from the average amount of money bet on the second, the fifth, and the sixth sequence (i.e. those sequences which make less likely to bet on Tail according to heuristic strategies). Thus, higher scores corresponded to high levels of susceptibility to the gambler's fallacy.

Procedure

Following methods of other studies (e.g., Panno, Lauriola, & Figner, 2013; Panno, Pierro, & Lauriola, 2013), participants were tested at two separate sessions, which were framed as two unrelated studies. In the first session, they completed the PRQ and the BSSS. The measures were administered in the classroom by professional trained experimenters. In the second session, each participant completed the *Non-Gambling* and the *Gambling Task* in a paper and pencil form.

Results

In line with the hypothesis, results showed that adolescents referred to probabilistic reasoning ability only in the *Non-Gambling Task*. Specifically, probabilistic reasoning ability was significantly and negatively related to the gambler's fallacy in the *Non-Gambling Task*, indicating that adolescents' performance was related to their reasoning abilities in the non-risky situation. On the contrary, the ability to reason with probability was no more significantly related to the performance in the *Gambling Task*, in which the susceptibility to the gambler's fallacy was significantly and positively related only to sensation seeking (Table 2).

Gambler's fallacy	Probabilistic reasoning	Sensation seeking
Non-Gambling Task	-.23*	.11
Gambling Task	.02	.22*
M (*SD*)	10.74 (*3.16*)	26.71 (*5.70*)

Table 2. Correlations between probabilistic reasoning ability, sensation seeking, and gambler's fallacy in the *Non-Gambling* and *Gambling Tasks*. (*p<.05)

Discussion

This study shows that probabilistic reasoning ability characterized adolescents' decision making only in the absence of risky conditions. This result suggested that without a potential winning money, only the cognitive system was activated. On the contrary, once the situation became risky, i.e. it changes from a 'neutral' to an 'affective-charged' context by the introduction of a potential winning money, only sensation seeking influenced adolescents' performance, thus suggesting that the affective system overcame the cognitive one in this kind of situation.

In terms of dual-system models, findings indicate that adolescents seem to use deliberating processes in the absence of affective involvement, but they do not seem to control affective impulses in situations of affective involvement. This finding is in line with what stated by the *fuzzy-trace theory* (Chick & Reyna, 2012; Reyna & Brainerd, 1995; Rivers, Reyna, & Mills, 2008), that assumes that risk taking in adolescence can be caused by a maturational lack of inhibition, particularly in situations of heightened emotional arousal. Moreover, our findings are also consistent with the *revised imbalance model* (Casey et al., 2008; Casey, Hare, & Galván, 2011; Galván, 2012), that claims that if it is true that adolescents are quite capable of rational decisions (Reyna & Farley, 2006), in emotionally salient situations, the affective system overcomes the cognitive system. Further, the neural imbalance typical of adolescence also interact with the context in which potentially risky decisions can be made: risky choices are largely influenced by the context in which they are presented (Galván, 2012). Indeed, individuals, especially adolescents, engage in risky choices because the context offer them a reward (Galván, 2012). It has been shown that neural systems that underlie reward are also those that precede risky decision making (Kuhnen & Knutson, 2005; Matthews, Simmons, Lane, & Paulus, 2004), suggesting a neural link between reward sensitivity and risk taking.

Study 2

The aim of the Study 2 was to explore the role of the cognitive system and the affective system in an *explicit* risky situation, i.e. a gambling condition with preliminary information given to participants. More specifically, we wanted to explore whether the role of probabilistic reasoning ability on adolescents' decision making changed between a risky context characterized by *high*-emotional arousal and a risky context with *low*-emotional arousal. In order to design these two different situations, a feedback after each choice was provided in the *high*-emotional arousal situation, while no feedback was provided in the *low*-emotional arousal situation. We based this decision on Figner et al. (2009)'s results in investigating risk taking in adolescents using the *Columbia Card Task* (CCT). Providing feedback (the "hot" CCT) was found to trigger more affective decision making, while receiving any feedback about the result of the decision until the end of the session (the "cold" CCT) resulted to trigger predominantly deliberative information processing.

Our first hypothesis was that adolescents would use probabilistic reasoning ability in both the risky situations. Indeed, from Brand and colleagues' research on decision making under conditions of risk (e.g., Brand et al., 2008; Brand et al., 2009; Schiebener et al., 2011), we know that the decision making performance in an explicit risky context is linked to cognitive functions such as logical ability, executive functions (i.e. categorization, set shifting, and rule learning), and knowledge of probabilities. Our second hypothesis was that sensation seeking would only affect adolescents' decision making in the high-emotional arousal condition, i.e. when feedback was provided.

Concerning the decision making task, in order to test our hypothesis in a context of explicit decision making under risk, we used the *Game of Dice Task* (GDT, Brand, Fujiwara et al., 2005) (Figure 1), a neuropsychological task in which individuals have to decide among different alternatives that are explicitly linked to a specific amount of gain or loss and have obvious winning probabilities that are stable over time. Therefore, individuals have the chance to calculate the risk associated with each alternative from the very beginning of the task. As such, probabilistic reasoning ability should help in recognizing which option can be more likely and in reasoning about the likelihood of the different options. Disadvantageous performances on this task have been associated with impaired decision-making process in clinical populations, such as adult pathological gamblers (e.g., Brand, Kalbe, et al., 2005), adolescent pathological Internet gamers (Pawlikowski & Brand, 2011), and adolescents with attention-deficit/hyperactivity disorder (Drechsler, Rizzo, & Steinhausen, 2008).

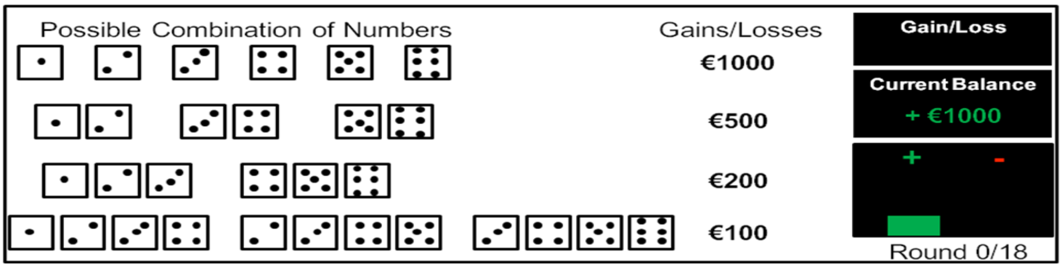


Figure 1. The Game of Dice Task (GDT, Brand, Fujiwara, et al., 2005).

Method

Participants

Participants were 201 adolescents (61% males, mean age=16.5 years, *SD*=1.60) who attended high school in Italy (Tuscany). Written informed assent was provided by students and written informed consent was provided by the parents if the student was a minor.

Instruments

Measures. As in the first Study, probabilistic reasoning was measured with the PRQ (Primi et al., 2014, and sensation seeking was revealed through the Italian version of the BSSS (Primi et al., 2011).

Task. In order to compare high- and low-emotional arousal situations, two versions of the same computerized task were administered. The *Game of Dice Task* – original version (GDT, Brand, Fujiwara, et al., 2005) was used to measure risky decision making in high-emotional arousal situation. Before beginning this task, participants were instructed to maximize their fictitious starting capital of €1,000 within 18 throws of a single virtual die. They were explicitly briefed about the rules of the game and the amounts of money associated with each of the possible options. In each trial, before the die is thrown, participants must bet on the outcome of the die throw choosing among one of the given options: a single number, or combinations of two, three, or four numbers. If they choose one of the six possible single number options (from "1" to "6", winning probabilit*y* = 0.17), they receive a fictitious gain of €1,000 when the chosen number is thrown but a fictitious loss of €1000 when one of the five other numbers not chosen is thrown. Choosing one of the three possible combinations of two numbers ("1,2" – "3,4" – "5,6", winning probabilit*y* = 0.33) is linked to a gain of €500 when one of the numbers included in the chosen combination is thrown, but a loss of €500 when one of the numbers not included in those combinations is thrown. A further alternative is to choose one of the two possible combinations of three numbers ("1,2,3" – "4,5,6", winning probabilit*y* = 0.50) linked to a potential gain/loss of €200. Finally, participants may choose one of the three possible combinations of four numbers ("1,2,3,4" – "2,3,4,5" – "3,4,5,6", winning probabilit*y* = 0.67) that will lead to a gain of €100 in the event that one of the four numbers chosen is thrown, but a loss of €100 when one of the numbers included in the other two combinations not chosen is thrown. In total, the participants can choose each of the 14 different alternatives (clustered in four groups) in each trial. The winning probabilities and amounts of gains and losses associated with each alternative remain stable during the entire task. After each throw, the gain or loss in money is indicated on the screen accompanied by a distinct sound (the jingle of a cash machine for a gain; a dull tone for a loss). The current total capital and number of remaining rounds are also displayed on the computer screen. In line with other studies (e.g., Brand, Kalbe et al., 2005; Brand, Labudda, & Markowitsch, 2006; Brand & Schiebener, 2013; Starcke, Tuschen-Caffier, Markowitsch, & Brand, 2010), the frequency of choosing the *risk-disadvantageous* or *high risky* choices (one single number and the two number combinations: winning probability of 0.33 or lower resulting in frequent and high losses in the long run) was calculated as a measure of risk taking. Thus, the higher the score, the more disadvantageous was the performance in terms of decision making.

The *Game of Dice Task* – Modified version (GDT-MOD, Brand, 2008) was used to measure risky decision making in low-emotional arousal situation. The modified GDT and the original GDT are similar with the exception that all feedback associated features have been removed from the modified version. This means that the participants cannot see the number thrown in each trial, and gains/losses are not indicated. Additionally, the current monetary balance is fixed to the starting capital (€1,000), and the color bars do not show whether a loss or gain has occurred. Participants also do not hear a tone for gains or losses. The participants see the screen of the original task on a notebook (fixed to the pattern at the beginning of the task) and are instructed in the same way as in the original GDT. They are told that their goal in the game is to win as much money as possible and to lose as little money as possible. In addition, all alternatives and gain/loss contingencies are explicitly explained and are indicated on the screen throughout the entire task (as in the original GDT). In the modified GDT, participants are told that the dice throws will not be seen on the screen and that they will not be informed about the outcome of each throw. They are also told that the computer will save the responses and that an overall feedback (final balance) will be provided at the end of the game. As in the GDT, the number of risky choices (i.e. one or two combinations choices) was calculated as a measure of risk taking.

Procedure

In line with Study 1, participants were tested at two separate sessions, which were framed as two unrelated studies. In the first session, participants completed the PRQ and the BSSS. The measures were administered in the classroom by professional trained experimenters. In the second session, each participant completed the decision making task in an individual setting on a desktop computer. Participants were randomly assigned to two groups. One group (n=102, 60% males, mean age=16.6 years, *SD*=1.53) was administered the GDT (Brand, Fujiwara, et al., 2005), in which feedback was provided, while the other group (n=99, 62% males, mean age=16.7 years, *SD*=1.58) was administered the GDT-MOD (Brand, 2008), in which feedback was not provided.

Results

In line with our hypotheses, results showed that adolescents referred to probabilistic reasoning ability in both the versions of the GDT. Specifically, we found that probabilistic reasoning was significantly and negatively related to the frequency of risky choices in the situation with feedback as in the situation without feedback, indicating that adolescents' decision making was related to reasoning abilities regardless of feedback. On the contrary, risky choices were significantly and positively related to sensation seeking only in the situation with feedback, i.e. only in conditions of high-emotional arousal (Table 3).

Risky choices	Probabilistic reasoning	Sensation seeking
GDT (With feedback)	-.24*	.20*
M *(SD)*	12.71 (*2.37*)	25.26 (*5.95*)
GDT-MOD (Without feedback)	-.37**	.06
M (*SD*)	12.69 (*2.17*)	23.42 (*5.57*)

Table 3. Correlations between probabilistic reasoning ability, sensation seeking, and the frequency of risky choices in the GDT and in the GDT-MOD. (*p<.05 **p<.01)

Results also showed that the frequency of risky choices was significantly lower in the GDT (*M*=4.10, *SD*=4.25) in comparison to the GDT-MOD (*M*=5.78, *SD*=4.71) ($t(199)=-2.66$, $p<.01$, Cohen's $d=.37$).

Discussion

Probabilistic reasoning characterized adolescents' decision making in explicit risky conditions, regardless of feedback. This finding suggests that the cognitive system was activated in this type of risky situation, in which information about the task contingencies are given to participants from the beginning of the task. Only in condition of high-emotional-arousal, adolescents' decision making was related to the affective system. On the contrary, in condition of low-emotional arousal, the affective system did not affect the decision making process. This is consistent with what claimed by Figner et al. (2009), according to which risky choice scenarios, such as choices between monetary lotteries without outcome feedback, typically trigger only minor affective processes.

The fact that adolescents' probabilistic reasoning has been found to be related to explicit decision making under risk provides evidence of the importance of probabilistic reasoning in decision making. Specifically, some Authors have said that if probabilities are not well understood it is likely that the choices in a risky situation will be suboptimal or too risk-seeking (Figner et al., 2009; Lauriola, Panno, Levin, & Lejuez, 2014). Among adolescents, proportional reasoning skill, a prerequisite skill for correct probabilistic reasoning, is considered as a fundamental ability for recognizing the most advantageous choice options in decision making tasks (Huizenga, Crone, & Jansen, 2007; Van Duijvenvoorde, Jansen, Visser, & Huizenga, 2010). Also, probability estimation, which entails deciding which choice has the largest chance of resulting in reward, is an important aspect of risk perception in the decision making process (Van Leijenhorst, Westenberg, & Crone, 2008). Finally, probabilistic reasoning was found to mediate the relationship between fluid intelligence and advantageous decision making in the GDT performance among adolescents (Donati, Panno, Chiesi, & Primi, 2014).

In terms of *dual-system models*, this study shows that adolescents seem to use deliberating process in situations of explicit risk. This finding is in line with what found by several studies conducted with adults showing the involvement of cognition in decision making under explicit risk (e.g., Brand et al., 2005; Brand et al., 2009; Schiebener et al., 2011). Findings also showed that explicit risky condition seem to lead adolescents to control affective impulses, since the cognitive system is activated even in the presence of feedback and it does not seem to be overcame by the affective system in such emotional arousal condition.

Finally, adolescents were found to choose more frequently the risky alternatives in the task version without feedback. This finding is in line with the study of Brand and colleagues (2009) conducted with healthy adults and revealing that participants who performed the original GDT showed a less disadvantageous decision making performance than those who performed the modified GDT. Consistent with that result, our study confirms that even in adolescents, processing feedback from previous trials may improve performance in a decision-making situation with explicit and stable rules for gains and losses.

General Discussion

This work offers empirical evidence that although adolescents' decision making was solely linked to the cognitive system in the non-risky situation, the affective system overcomes the cognitive system when the decision making situation becomes risky, i.e. when a potential winning money was introduced. Additionally, it suggests that providing information about the task might interfere with the imbalance between the two systems that usually verifies in the presence of risky situations. This would explain why adolescents' decision making in the condition with feedback was related not only to the affective, but also to the cognitive system, while, without feedback, adolescents' performance was solely related to the cognitive system. Moreover, it indicates that providing feedback was associated with less risk taking than not providing it.

From a general standpoint, the present work shows that adolescent decision making seems to be impaired in situations of implicit risk taking, and this impairment seems to be related to a difficulty in inhibiting affective impulses. However, when the risky situation becomes more explicit, i.e. information about the task contingencies are given, adolescents use their cognitive system; moreover, this system seems to be activated even in the situation of heightened emotional arousal, i.e. when a feedback after each choice was provided.

Findings underline the importance of the context in interacting with the imbalance between the cognitive and the affective system in influencing adolescents' decision making in risky situations (Galván, 2012). Indeed, when the level of information available on the decision-making task became more explicit for adolescents, they showed to refer to their cognitive abilities in terms of probabilistic reasoning skills.

From the educational point of view, the most important finding of this work was that providing explicit information on pros and cons of choice options can lead adolescents to use the cognitive system. However, a cognitive decision making can occur only in situation of low-emotional arousal. Then, preventive interventions designed to reduce adolescents' risk-taking behaviors should be addressed to educate adolescents to approach risk-taking contexts as much as possible as reasoning contexts. In other words, it would be interesting to develop educational interventions aimed not only at promoting the development of adolescents' understanding of probability, for example by providing them real-world experiences with random generators (Morsanyi, Handley, & Serpell, 2012), but also at fostering their tendency to understand all the elements in everyday decision making situations that can be read as information on pros and cons of the different alternatives among which each time they have to decide. This might help them in using their reasoning abilities regardless of risky context.

References

Brand, M. (2008). Does the feedback from previous trials influence current decisions? A study on the role of feedback processing in making decisions under explicit risk conditions. *Journal of Neuropsychology*, *2*, 431-443.

Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J., & Markowitsch, H. J. (2005). Decision making deficits of Korsakoff patients in a new gambling task with explicit rules: Associations with executive functions. *Neuropsychology*, *19*, 267–277. doi: 10.1037/0894-4105.19.3.267.

Brand, M., Heinze, K., Labudda, K., & Markowitsch, H. J. (2008). The role of strategies in deciding advantageously in ambiguous and risky situations. *Cognitive Processing*, *9*, 159–173. doi: 10.1007/s10339-008-0204-4.

Brand, M., Kalbe, E., Labudda, K., Fujiwara, E., Kessler, J., & Markowitsch, H. J. (2005). Decision-making impairments in patients with pathological gambling. *Psychiatry Research*, *133*, 91–99. doi: 10.1016/j.psychres.2004.10.003.

Brand, M., Laier, C., Pawlikowski, M., & Markowitsch, H. J. (2009). Decision making with and without feedback: The role of intelligence, strategies, executive functions, and cognitive styles. *Journal of Clinical and Experimental Neuropsychology*, *31*, 984–998. doi: 10.1080/13803390902776860.

Brand, M. & Schiebener, J. (2013). Interactions of age and cognitive functions in predicting decision making under risky conditions over the life span. *Journal of Clinical and Experimental Neuropsychology*, *35*(1), 9-23. doi: 10.1080/13803395.2012.740000.

Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences, 1124*, 111-126. doi: 10.1196/annals.1440.010.

Chein, J., Albert, D., O'Brien, L., Uckert, K., & Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Developmental Science, 14*, F1-F10. doi: 10.1111/j.1467-7687.2010.01035.x

Chick, C. F., & Reyna, V. F. (2012). A fuzzy trace theory of adolescent risk taking: Beyond self-control and sensation seeking. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (2012). T*he Adolescent Brain: Learning, Reasoning, and Decision Making.* Washington, DC, USA: American Psychological Association.

Cohen, J. D. (2005). The vulcanization of the human brain: A neural perspective on interactions between cognition and emotion. *Journal of Economic Perspectives, 19,* 3-24.

Delfabbro, P., Lahn, J., & Grabosky, P. (2006). It's not what you know, but how you use it:

Statistical knowledge and adolescent problem gambling. *Journal of Gambling Studies*, *22*, 179-193. doi: 10.1007/s10899-006-9009-5.

Delfabbro, P., Lambos, C., King, D., & Puglies, S. (2009). Knowledge and beliefs about gambling in Australian secondary school students and their implications for education strategies. *Journal of Gambling Studies, 25*, 523-539. doi: 10.1007/s10899-009-9141-0.

Donati. M. A., Chiesi, F., & Primi, C. (2013). A Model to Explain At Risk/Problem Gambling among Male and Female Adolescents: Gender Similarities and Differences. *Journal of Adolescence*, *36*, 129-137. doi: http://dx.doi.org/10.1016/j.adolescence.2012.10.001.

Donati, M. A., Panno, A., Chiesi, F., & Primi, C. (2014). A mediation model to explain decision making under conditions of risk among adolescents: The role of fluid intelligence and probabilistic reasoning. *Journal of clinical and experimental neuropsychology*, *36*(6), 588-595.

Drechsler, R., Rizzo, P., & Steinhausen, H. C. (2008). Decision-making on an explicit risk taking task in preadolescents with attention-deficit/hyperactivity disorder. *Journal of Neural Transmission*, *115*, 201-209. doi: 10.1007/s00702-007-0814-5.

Figner, B., Mackinlay, R., Wilkening, F., & Weber, E. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 709–730. doi: http://dx.10.1037/a0014983.

Galván, A. (2012). Risky behavior in adolescents: The role of the developing brain. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (2012). T*he Adolescent Brain: Learning, Reasoning, and Decision Making.* Washington, DC, USA: American Psychological Association.

Galván, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., et al. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. Journal of Neuroscience, 26, 6885-6892.

Galván, A., Hare, T. A., Voss, H., Glover, G., & Casey, B. J. (2007). Risk-taking and the adolescent brain: Who is at risk? Developmental Science, 10, F8–F14.

Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. Developmental Psychology, 41, 625–635.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Heuristics and biases: The psychology of intuitive judgment. Cambridge, UK: Cambridge University Press.

Gupta, R., & Derevensky, J. L. (1998). An empirical examination of Jacobs' General Theory of Addiction: do adolescents gamblers fit the theory? Journal of Gambling Studies, 14, 17-49.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. Personality and Individual Differences, 32, 401-414.

Huizenga, H. M., Crone, E. A., & Jansen, B. J. (2007). Decision-making in healthy children, adolescents and adults explained by the use of increasingly complex proportional reasoning rules. *Developmental Science*, *10*(6), 814-825. doi: 10.1111/j.1467-7687.2007.00621.x.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. Cambridge, UK: Cambridge University Press.

Keating, D. (2004). Cognitive and brain development. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 45-84). New York: Wiley.

Knoch, D., & Fehr, E. (2007). Resisting the power of temptations: The right prefrontal cortex and self-control. *Annals of the New York Academy of Sciences, 1104,* 123-134.

Kuhn, D. (2006). Do cognitive changes accompany developments in the adolescent brain? *Perspectives on Psychological Science, 1,* 59-67.

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, *47*(5), 763-

770.

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual Differences in Risky Decision Making: A Meta–analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task. *Journal of Behavioral Decision Making, 27*, 20-36. doi: 10.1002/bdm.1784.

Matthews, S. C., Simmons, A. N., Lane, S. D., & Paulus, M. P. (2004). Selective activation of the nucleus accumbens during risk-taking decision making. *Neuroreport*, *15*(13), 2123-2127.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004, October 15). Separate neural systems value immediate and delayed monetary rewards. *Science, 306,* 503-507.

Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. (2009). The effects and side-effects of statistics education: Psychology students'(mis-) conceptions of probability. *Contemporary Educational Psychology*, *34*(3), 210-220.

Morsanyi, K., Handley, S. J., & Serpell, S. (2012). Making heads or tails of probability: An experiment with random generators. *British Journal of Educational Psychology*, *83*(3), 379-395. doi: http://dx.10.1111/j.2044-8279.2012.02067.x.

Nower, L., Derevensky, J. L., & Gupta, R. (2004). The relationship of impulsivity, sensation seeking, coping and substance use in youth gamblers. P*sychology of Addictive Behaviours*, *18,* 49-55. http://dx.doi.org/10.1037/0893-164X.18.1.49.

Panno, A., Lauriola, M., & Figner, B. (2013). Emotion regulation and risk taking: Predicting risky choice in deliberative decision making. *Cognition & Emotion*, *27*, 326-334. doi: doi.org/10.1080/02699931.2012.707642.

Panno, A., Pierro, & Lauriola, M. (2013). Self-regulation predicts risk taking through people's time horizon. *International Journal of Psychology*. doi: http://dx.10.1002/ijop.12026.

Pawlikowski, M., & Brand, M. (2011). Excessive Internet gaming and decision making: Do excessive World of Warcraft players have problems in decision making under risky conditions? *Psychiatry Research*, *188*, 428-433. doi: 10.1016/j.psychres.2011.05.017.

Primi, C., Morsanyi, K., & Chiesi, F.,(2014). Measuring the basics of probabilistic reasoning: the IRT-based construction of the probabilistic reasoning questionnaire. *Proceedings of the 9th International Conference on Teaching Statistics* (ICOTS), Flagstaff, Arizona, USA.

Primi, C., Narducci, R., Benedetti, D., Donati, M. A., & Chiesi, F. (2011). Validity and reliability of the Italian version of the Brief Sensation Seeking Scale (BSSS) and its invariance across age and gender. Testing, Psychometrics, Methodology in Applied Psychology, TPM, 18, 1-11.

Reyna, V. F., Chapman, S. B., Dougherty, M. R., & Confrey, J. (2012). T*he Adolescent Brain: Learning, Reasoning, and Decision Making.* Washington, DC, USA: American Psychological Association.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1-75.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making implications for theory, practice, and public policy. *Psychological Science In The Public Interest, 7*, 1-44.

Rivers, S. E., Reyna, V. F., & Mills, B. (2008). Risk taking under the influence: A fuzzy-trace theory of emotion in adolescence. *Developmental Review*, *28*(1), 107-144.

Reyna, V. F., & Rivers, S. E. (2008). Current theories of risk and rational decision making. *Developmental Review: DR*, *28*(1), 1.

Schiebener, J., Zamarian, L., Delazer, M., & Brand, M. (2011). Executive functions, categorization of probabilities, and learning from feedback: What does really matter for decision making under explicit risk conditions? *Journal of Clinical and Experimental Neuropsychology*, *33*(9), 1025–1 039. doi: http://dx.doi.org/10.1080/13803395.2011.595702.

Scholes-Balog, K. E., Hemphill, S. A., Dowling, N. A., & Toumbourou, J. W. (2014). A prospective study of adolescent risk and protective factors for problem gambling among young adults. *Journal of Adolescence*, *37*(2), 215-224. doi:

http://dx.doi.org/10.1016/j.adolescence.2013.12.006

Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in cognitive sciences*, *15*(1), 11-19.

Skoukaskas, N., & Satkeviciute, R. (2007). Adolescent pathological gambling in Kaunas, Lithuania. *Nordic Journal of Psychiatry*, *61*, 86-91.

Somerville, L. H., Jones, R. M., & Casey, B. J. (2010). A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition, 72*, 124-133. doi: 10.1016/j.bandc.2009.07.003.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28,* 78-106.

Turner, N. E., MacDonald, J., Bartoshuk, M., & Zangeneh, M. (2008). Adolescent gambling behavior, attitudes, and gambling problems. I*nternational Journal of Mental Health and Addiction*, *6*, 223-237. http://dx.doi.org/10.1007/s11469-007-9117-1

Turner, N. E., Zangeneh, M., & Littman-Sharp, N. (2006). The experience of gambling and its role in problem gambling. *International Gambling Studies*, *6*, 237-266.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, *76*(2), 105.

Van Duijvenvoorde, A. C. K., Jansen, B. R. J., Visser, I., & Huizenga, H. (2010). Affective and cognitive decision-making in adolescents. *Developmental Neuropsychology, 35,* 539–554. doi: http://dx.10.1080/87565641.2010.494749.

Van Leijenhorst, L., Westenberg, P. M., & Crone, E. A. (2008). A developmental study of risky decisions on the Cake Gambling Task: Age and gender analyses of probability estimation and reward evaluation. *Developmental Neuropsychology*, *33*(2), 179-196. doi: http://dx.10.1080/87565640701884287.

Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111,* 430-445.

Levels of Reasoning of Middle School Students about Data Dispersion in Risk Contexts

Ernesto Sánchez
Departamento de Matemática Educativa, Cinvestav-IPN, México
Antonio Orta
Escuela Nacional para Maestras de Jardines de Niños, México

Abstract: The aim of this research study is to explore students' reasoning concerning variation when they compare groups and have to interpret dispersion in terms of risk. In particular, we analyze in this paper the responses to two problems from a questionnaire administered to 82 ninth-grade students. The problems consist of choosing between two and three groups of data by comparing them. The first one composed of losses and winnings coming from a hypothetical game; the second is about medical treatments. The results show the difficulty students had in interpreting variation in a risk context. Although they identify the data group with more variation, this is not enough for interpreting the variation in terms of risk and making a rational decision. The psychological categories of risk-seeking and risk-aversion are used to explain the behavior of students who choose one group or another when they identify correctly the risk in each situation. As a conclusion, it is suggested that more risk context situations should be studied.

Keywords: variation, dispersion, variability, risk, middle school students.

Introduction

Variation is the underlying reason for the existence of statistics (Watson, 2006, p. 217) since variability is in everywhere and therefore in data. Moore (1990) emphasized the omnipresence of variation and the importance of modeling and measure variation in statistics; Wild & Pfannkuch (1999) included the perception of variation as part of the fundamental types of statistical reasoning. Burrill & Biehler (2011) proposed a list of seven fundamental statistical ideas in which variation is the second after data. But this growing recognition of the importance of variation in basic education is relatively recent. Just 18 years ago Shaughnessy (1997) wondered where the educational research on variability was and called to the community of statistical educators to investigate on statistical variation. As a consequence several researchers began to explore scenarios that would allow students to display their understanding about variation.

As a fundamental idea it is supposed that variation "can be taught effectively in some intellectually honest form to any child at any stage of development" (Heitele, 1975, p. 187), so research to improve instruction in variation can be located at any level. In a growing number of studios on students' thinking about variation, contexts and problems have been proposed to encourage students from different scholar levels to perceive, describe or/and measure variation in data. For example, variability in sampling (Watson & Moritz, 2000), chance (Watson & Kelly, 2004), repeated measurement, and natural variation in the plants growth (Lehrer & Schauble, 2007, Petrosino, Lehrer &Schauble, 2003), weather (Reading, 2004), all these situations have been explored for develop statistical thinking of students at different levels. Risk situation could provide another context to investigate variability. In this paper we propose a way to formulate some decision making problems in risk contexts where dispersion is relevant and we explore the reasoning of middle school students in front of these problems.

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 275-289

Conceptual framework

The concepts that constitute our conceptual framework try to provide a comprehensive understanding of the study. The first is that of tasks as a key element in teaching and learning. In statistics the context of the tasks has an important role for understanding statistical ideas; therefore the second concept of our framework is risk context in its relation to variation. Reasoning as a way to close to students' thinking is our third concept. Finally, we include SOLO model as an instrument to organize the reasoning of students.

Tasks

An important part of research in mathematics and statistics education is to seek high-level tasks that promote the capacity to think, reason, and solve problems related to the fundamental ideas of the study area. The tasks should also encourage to the students to engage with the concept to be learned. In statistical unlike mathematics, reasoning must articulate abstract mathematical ideas with real situations through the data, i.e. statistical reasoning is intimately linked with the contexts. The formulation of tasks that need to collect data is not sufficient to emerge themselves abstract ideas (concepts) that must be learned. Nor is it enough to organize students to work with tasks that only require the manipulation of concepts and properties independently to situations. Finding a balance between these extremes depends largely on the choice of good problems. These are the means by which the actions of the teacher are transformed in student engagement in reflection and action.

Problems on decision making under uncertainty are common in statistics; this kind of problems has been widely used to promote, and also to analyze, some important aspects of the statistical reasoning of people. On the other side, tasks on comparing two groups are frequently used to engage students in reasoning about data since in many statistical studies is necessary compare groups. We are proposing two problems on decision making where comparing groups of data is required and dispersion is significant, in addition their solution implies some risk preferences.

Risk context

The interpretation of dispersion depends on the situation from which the data come. One kind of elemental problems where variation could emerge can be formulated in risk context. When the uncertainty present in a process implies any threat to the effect of a result, it is called a risk. These situations appear when there are potential and unwanted results that, as a consequence, lead to losses or damages. Defining risk means to specify both the valuable and unwanted results in a way that reflects the value attributed to them. Analyses of risk situations offer information for decision-making. The theory of decision making under situations of risk has two aspects. On one hand, it defines abstract rules for what people should do; and, on the other, it studies what people really do when facing the risk.

The theory of decision making in situations of risk is a broad theory, but what is necessary for our research are very basic elements. These are restricted to the concepts of prospect, risk aversion and risk seeking as they have been defined by Kahneman and Tversky (2000).

Consider the following problem:

Problem: Choose between	
A: p_0 chance to win X_0	B: p_1 chance to win X_1
q_0 chance to win Y_0	q_1 chance to win Y_1

The overall utility of each game is:

$$U_A = p_0 \times X_0 + q_0 \times Y_0$$
$$U_B = p_1 \times X_1 + q_1 \times Y_1$$

According to classical theory of utility, comparing U_A to U_B the most suitable game is determined.

These decision making problems are generalized as a choice between prospects or gambles in the following way. A prospect $(x_1, p_1; x_2, p_2; \dots; x_n, p_n)$ is a contract that yields outcome x_i with probabilityp_i, where$p_1 + p_2 + \cdots + p_n = 1$. In Prospect Theory (Kahneman & Tversky, 2000) a utility function for outcomes $u(x_i)$ is postulated. In our case we suppose that such function is the identity $u(x_i) = x_i$ since the outcomes are given in the monetary value that the gamblers win (or in time that the patients live with each treatment). Then the overall utility of the prospect is the expectation:

$$U(x_1, p_1; \dots; x_n, p_n) = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$$

Consider a situation where a sample $x_1, x_2 \dots; x_n$of outcomes of a game is given and their corresponding probabilities are unknown. Due to lack of information may be reasonable to assume that the probability of each outcome is 1/n. In such case the overall utility is the arithmetical mean:

$$U\left(x_1, \frac{1}{n}; \dots; x_n, \frac{1}{n}\right) = \frac{x_1}{n} + \cdots + \frac{x_n}{n} = \bar{x}$$

Now it is possible to formulate problems of decision-making as follows:

Problem. The gains of realizations of *n* times the game A and *m* the game B are:

Game A: $x_1, x_2 \dots; x_n$

Game B: $y_1, y_2 \dots; y_m$

Which of the two games would you choose to play in?

If the games are thought as prospects, according to classical theory of utility, solution is obtained by comparing $\bar{x}$ and$\bar{y}$. (This procedure of solution is not easily understood by students).

However, when the expected gains coincide $\bar{x} = \bar{y}$ not necessarily the most appropriate decision is to choose any of the two games since the dispersion of the values in each set can be significant for the decision maker. Indeed, the comparison of the dispersion of each set gives an account of the difference in risk terms. Consider for example the following problem:

Problem. The gains of 2 realizations of two games are:

Game A: $-1, 3$

Game B: $1,\ 1$

Which of the two games would you choose to play in?

The overall utility of each game is 1 but for people the games are not equivalent if games are thought as prospects:

A:	50% chance to win 3 50% chance to loss 1	B:	1 for sure

Frequently people prefer game B because they do not like to take risks. In the research of how people answer this kind of problems psychologists have elaborated the concepts of risk aversion and risk seeking:

> The preference for a sure gain is an instance of risk aversion. In general, a preference for a sure outcome over a gamble that has higher or equal expectation is called risk aversion, and the rejection of a sure thing in favor of a gamble of lower or equal expectation is called risk seeking (Kahneman & Tversky, 2000, p. 2)

It is worth noting that in a game, the dispersion of gains (including losses) can be considered a measure of risk. With this idea, an extension of these attitudes toward risk can be developed to explain some of the behaviors observed in the responses of students to decision-making problems studied in this work. Consider a problem of making decision on two data sets whose arithmetic means is the same but with different dispersion, and suppose the data belong to a variable that carries risk. Let's say that a preference is motivated by risk aversion when an option whose data have less dispersion over another whose data have greater dispersion is preferred. The decision is motivated by risk seeking when the option whose data have greater dispersion is chosen.

Reasoning

The statistics education community has distinguished three overlapping areas of statistics to organize and analyze the objectives, activities and results of statistical learning: statistical literacy, reasoning and thinking. This study is located in the area of statistical reasoning. The purpose of the research on statistical reasoning is to understand how people reason with statistical ideas (Garfield & Ben-Zvi, 2008) in order to propose features to create learning scenarios. It is worth then, to make some remarks on the idea of reasoning in general. For this we take some lessons from *inferentialism* (Brandom, 2000); Bakker & Derry, (2011) point out three lessons:

- Concepts should be primarily understood in terms of their role in reasoning and inferences
- To understand concept use we should privilege holism over atomism
- Privileging an inferentialist approach to education over representationalist one.

The first implies that "to talk about concepts is to talk about roles in reasoning" (Brandom, 2000, p. 11); concepts are understood in social practices of asking and giving reasons.

The second lesson tells us that "One cannot have any concepts unless one has many concepts. For the content of each concept is articulated by its inferential relations to other concepts. Concepts, then, must come in packages." (Brandom, 2000, p. 15-16). The third lesson is the recommendation of Bakker and Derry for using an inferentialist approach for analyzing educational process.

When students try to justify their responses, elements that they think are important to the situation are revealed; in particular, the data they choose, operations made with these and knowledge and beliefs on which they rest for doing that, are important in reasoning. Unfortunately, the answers given by the students are usually not as explicit as to give us information on these three aspects; anyway, we will try to identify their main features of reasoning from their responses. For this task we will rely on our interpretation of the SOLO model.

SOLO Model

The Biggs and Collis (1982, 1991) *Structure of Observed Learning Outcomes* (SOLO) model has been used by many researchers to identify levels of student reasoning on different concepts. The SOLO model is based on the assumption that development can be represented in hierarchical structures. In a similar way to the stage theory of development of Piaget, five modes of representation are postulated in SOLO model: Sensorimotor (from birth), Ikonic (from around 18 months), Concrete symbolic (from around six years), Formal (from around 14 years) and Post-formal (from around 20 years). "Modes, then, are levels of abstraction, progressing from concrete actions to abstract concepts and principles, which form the basis of the developmental stages"(Biggs & Collis, 1991, p. 62).

Within each mode and related to a task in a conceptual web, levels of reasoning that progress from incompetence to expertise can be identified. At the *Prestructural* level (P) responses only show that students engage to task but do not use any relevant aspect to its solution. In the *Unistructural* level responses that have one relevant aspect to the task solution are classified. The responses at *Multistructural* level present more than one relevant aspect but without integrate them. For the *Relational* level responses integrate in a coherent way more than one relevant aspect to the task. Finally, at the Extended Abstract (EA) level responses would show a higher abstract response. A hierarchy that describes levels of increasing complexity in a students' reasoning is obtained.

Some relevant aspects of the solution to the task are chosen according to an a priori analysis, however, is in the process of analyzing data that emerge unanticipated patterns which reveal peculiar forms of reasoning of students. It should be clarified that in the data analysis, the responses of the students are classified in different levels but are not intended to classify students. Finally we subscribe the Watson commentary about that SOLO"[..] is a useful model because it stresses that it is observed in students' responses that is analyzed, not what the observer thinks the student might have meant" (Watson, 2006, p. 13).

Method

The participants were 82 students, the teacher in charge of both groups, and the researchers who are the authors of this work. Students (aged 14 to 16) belonging to two different ninth grade groups in a private school in Mexico City (In Mexico 9th grade is the last year of middle school). Two problems were designed to explore the reasoning of the students; the first is related to a game where the risk is represented by possible losses, the second is related to life time after follow different treatments.

Problem 1

In a fair, the attendees are invited to participate in one of two games, but not in both. In order to know which game to play, John observes, takes note and sorts the results of 10 people playing each game. The losses (-) or cash prizes (+)obtained by the 20 people are shown in the following lists:

Game 1:

15	-21	-4	50	-2	11	13	-25	16	-4

Game 2:

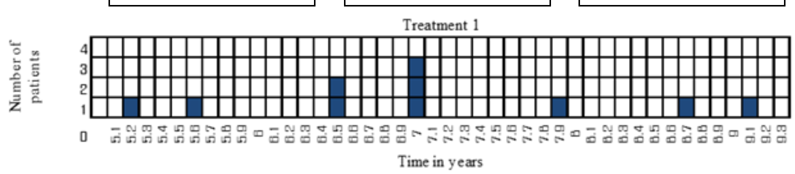
120	-120	60	-24	-21	133	-81	96	-132	18

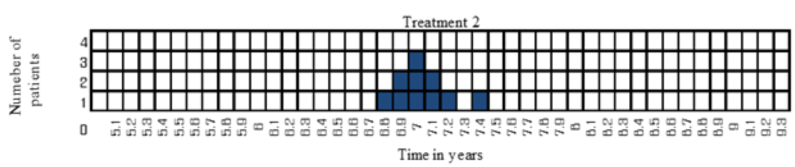
a) If you had the possibility of playing only one of the two games, which one would you choose? Why?
b) In which of the two games do the data have more variability? Why?

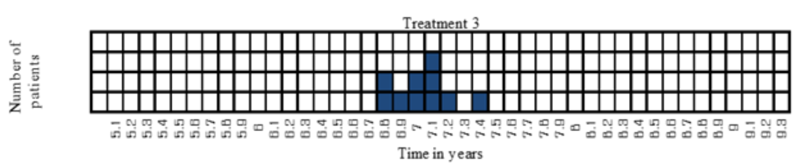
Problem 2

Consider you must advice a person who suffers from a severe, incurable and deathly illness, which may be treated with a drug that may extend the patient's life for several years. It is possible to choose between three different treatments. People show side effects to the medication; while in some cases the drugs have the desired results, in some others the effects may be more favorable or more adverse. The following lists show the number of years the patients have lived after being treated with one of the different options; each number in the list corresponds to the time in years a patient has survived with the respective treatment. The graphs corresponding to the treatments are shown after.

Time in years (Treatment 1)	Time in years (Treatment 2)	Time in years (Treatment 3)
5.2	6.8	6.8
5.6	6.9	6.8
6.5	6.9	6.9
6.5	7.0	7.0
7.0	7.0	7.0
7.0	7.0	7.1
7.0	7.1	7.1
7.8	7.1	7.1
8.7	7.2	7.2
9.1	7.4	7.4







a) What kind of treatment would you prefer (1, 2 ó 3)? Why?

b) In which of the treatments there is more variability? Explain your answer

An activity was designed and developed during two teaching sessions of 50 minutes each. The students solved some problems where they have to analyze data and calculate means, ranges and mean deviation but without learned more widely their meanings. The two problems were administered before and after the activities; however the analysis presented below is made taken together all data. Some remarks about the differences between frequencies of responses of pre-test and post-test are mentioned at the end.

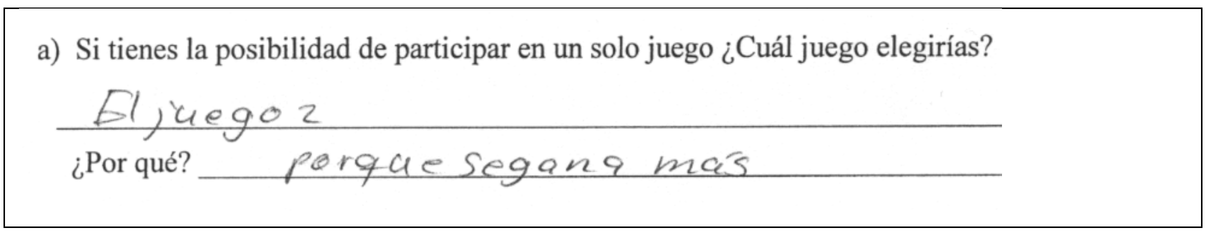
Results

In this section we present examples of responses to questions 1a, 1b, and 2a described above, in order to show the kind of answers that were classified at each SOLO level. At the end of the examples of each question a table with the frequencies of responses of each level is presented.

Examples of Responses to Question 1a

If you had the possibility of playing only one of the two games, which one would you choose? Why?

Prestructural level. In the Figure 1 response, the student chose game 2, "because you win more", but there is no evidence of how the data are used, although one can assume that the answer is motivated by something the student perceived in them. The answer is circular because the questions: What game would you choose?, and: In what game you win more?, are equivalent. This kind of responses provides no progress in understanding the situation.

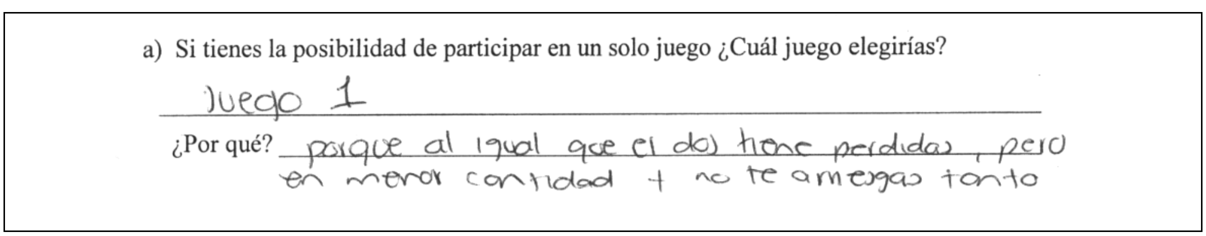

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

El juego 2

¿Por qué? porque se gana más

Figure 1. Prestructural level answer.

Unistructural level. In the response of the Figure 2, the student chooses Game 1 "because you can also lose as in game two but fewer and you risk less", that is he thinks that in game 2 he loses less and provides an indication that he compares the loss of both games (probably the minimum of each set or the sum of losses) and also alludes to risk. Indeed, it is possible that the way in which the student approaches the problem is influenced by risk aversion, since he skews his attention toward the losses, ignoring the information that provides positive gains.

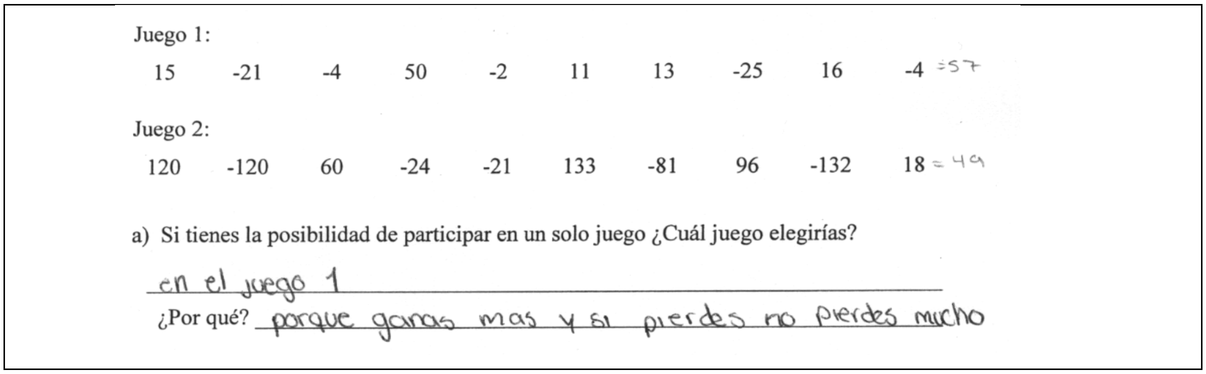

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

juego 1

¿Por qué? porque al igual que el dos tiene perdidas, pero en menor cantidad + no te arriesgas tanto

Figure 2. Unistructural level answer.

In the response of Figure 3, the student chose Game 1 arguing "because you win more and if you lose, do not lose too much". In addition he wrote in the margin of each data list the result of the sum of the values, but with an error in the first sum. Although we considered that responses, in which students add and compare the gains of each game, are Multi-structural, in this case, the answer is located in Uni-structural, taking into account that the student made a mistake in the sum.

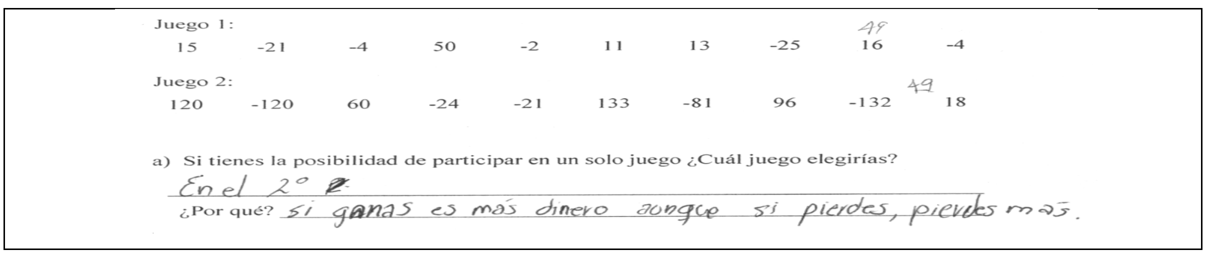

Juego 1:
15 -21 -4 50 -2 11 13 -25 16 -4 =57

Juego 2:
120 -120 60 -24 -21 133 -81 96 -132 18 = 49

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?
en el juego 1
¿Por qué? porque ganas mas y si pierdes no pierdes mucho

Figure 3. Unistructural level answer (mistake in the sum).

Multistructural level. In response of the Figure 4 the student chooses Game 2, and argues "If you win you get more money but if you lose, lose more". There are features in his worksheet showing that he added and compared the values of each data list. However, he does not draw the conclusion that both games yield the same profit, instead he put attention on the extreme values, since effectively in Game 2 it is possible win a lot, but also lose a lot. This response reflects risk seeking. The response is classified in Multiestructural because the student considered two relevant aspects of the problem: a pre-figuration of the arithmetic mean and the range, but without relate them in a convenient way.

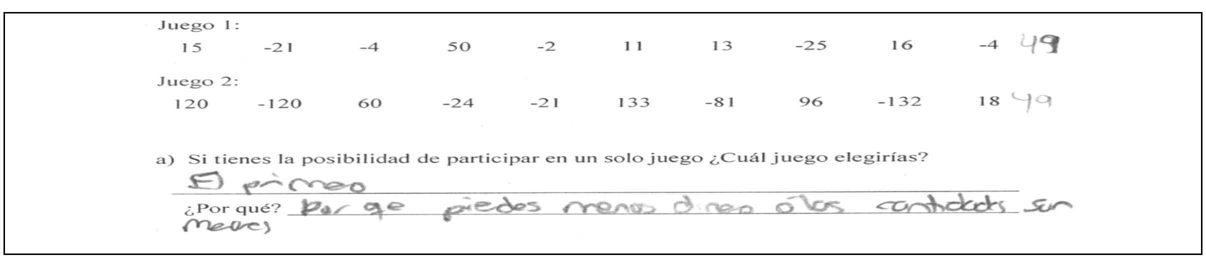

Juego 1:
15 -21 -4 50 -2 11 13 -25 16 -4 49

Juego 2:
120 -120 60 -24 -21 133 -81 96 -132 18 49

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?
En el 2°
¿Por qué? si ganas es más dinero aunque si pierdes, pierdes más.

Figure 4. Multistructural level answer (risk seeking).

The response of Figure 5, student chose the game 1, "because you lose less money or the quantities are lower". This response is similar to the previous because the student adds and realizes that the results of both list are equal, but like before, does not draw the conclusion that in both games, on average, the same profit is obtained. He also pays attention to extreme values, but now chooses Game 1, probably due to risk aversion.


Juego 1:
15 -21 -4 50 -2 11 13 -25 16 -4 49

Juego 2:
120 -120 60 -24 -21 133 -81 96 -132 18 49

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?
El primeo
¿Por qué? Porqe piedos menos dineo ó los contidads son meores

Figure 5. Multistructural level answer (risk aversion).

Relational level. There was no response classified in this level. They would be responses in which the students realize that both games have the same utility but are different respect to risk, and then they made a decision according his risk preferences.

Percentages by level:

Level	Pre-test frequency	Post-test frequency
Pre-structural	77 (89 %)	54 (70 %)
Uni-structural	6 (7 %)	4 (5 %)
Multi-structural	3 (3 %)	19 (25 %)
No response	1 (1 %)	
Total	87	77

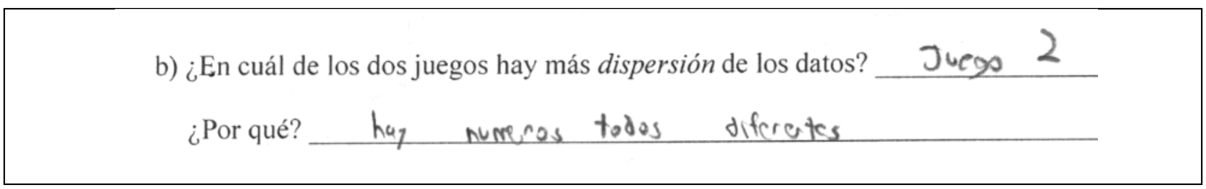
Table 1. Number of responses by level

With respect to Table 1, frequency of pre-structural responses decreased in post-test respect to pre-test while that of the multi-structural responses increased. The main reason is that there was greater use of the procedure for adding the amounts of each list, probably as a result of the students solved problems that used the arithmetic mean in the episodes of instruction. However, in the answers to post-test, students only added the amounts of each list without actually calculating the arithmetic mean.

Examples of Responses to Question 1b.

In which of the two games do the data have more variability? Why?

Prestructural level. We found answers in which the student makes a choice of a game but the argument offered would lead to the election of the other or such argument does not correspond with the choice made. We have classified these responses in pre-structural. For example, in Figure 6 the student chose the game 2 because "there are all different numbers"; however this is not true.

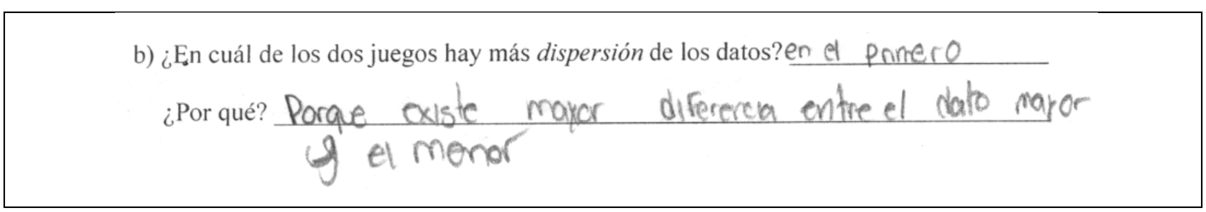


b) ¿En cuál de los dos juegos hay más *dispersión* de los datos? Juego 2

¿Por qué? hay numeros todos diferetes

Figure 6. Prestructural level answer.

In Figure 7 the student chooses Game 1 because "there is a greater difference [in the game 1] between the highest datum and lowest datum [than in the game 2]", but actually meets this property the game 2 and not 1. Other responses classified in prestructural give vague arguments as "because there are very low data and very high data" or "because there are very large quantities"

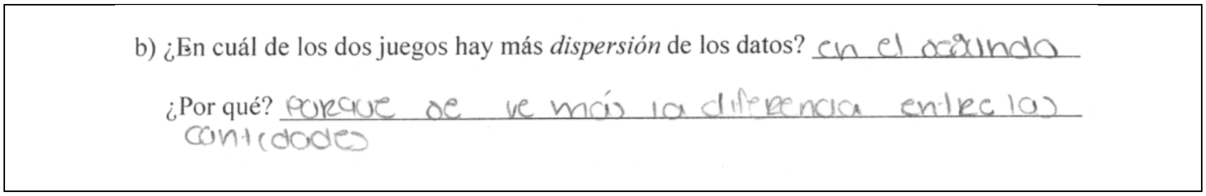


b) ¿En cuál de los dos juegos hay más *dispersión* de los datos? en el primero

¿Por qué? Porque existe mayor diferencia entre el dato mayor y el menor

Figure 7. Prestructural level answer.

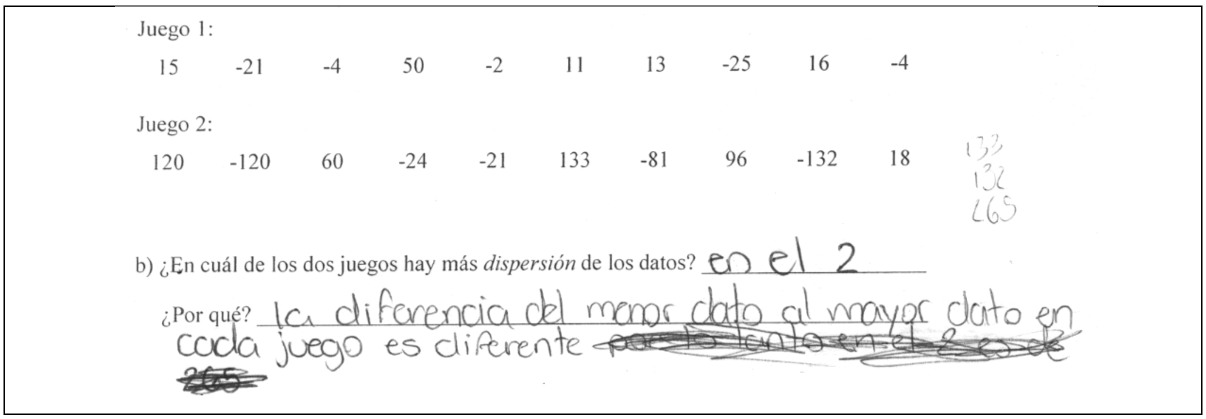
Unistructural level. In the Unistructural level we have classified responses, whose argument provides evidence that the student observed some trait in data related to dispersion and his choice is consistent. For example, in response of Figure 8, the student chooses Game 2, "because there are greater differences between the amounts". In this response the differences compared are not specified, but the fact of having chosen the game whose data are more scattered, indicates that probably the range of each game, or the sum of the differences between successive data, was considered.


b) ¿En cuál de los dos juegos hay más *dispersión* de los datos? en el segundo

¿Por qué? porque se ve más la diferencia entre las cantidades

Figure 8. Unistructural level answer.

Multistructural level. The responses in which game two was chosen as the most dispersed and the decision was argued by comparing the ranges of both data sets were classified as Multistructural. In Figure 9 the student chose the game 2 and argued "the difference between the maximum and minimum is different in each game" and in his worksheet wrote 133 + 132 = 265 indicating that he considered at least the range of game 2 data.


Juego 1:

15 -21 -4 50 -2 11 13 -25 16 -4

Juego 2:

120 -120 60 -24 -21 133 -81 96 -132 18

133
132
265

b) ¿En cuál de los dos juegos hay más *dispersión* de los datos? en el 2

¿Por qué? la diferencia del menor dato al mayor dato en cada juego es diferente ~~por lo tanto en el 2 es de 265~~

Figure 9. Multistructural level answer

Relational level. There is no response classified in this level.

Level	Pre-test frequency	Post-test frequency
Prestructural	74 (85 %)	48 (62 %)
Unistructural	9 (11 %)	14 (18 %)
Multistructural	2 (2 %)	13 (17 %)
No response	2 (2 %)	2 (3 %)
Total	87	77

Table 2. Frequencies by level

Examples of Responses to Question 2a.

What kind of treatment would you prefer (1, 2 or 3)? Why?

Prestructural level. The arguments of the pre-structural responses do not provide evidence that students have taken into account the data or how the data were used to make the decision. For example, in one response where treatment X was chosen, the student argues "there is more time and quality of life". In a class of responses where treatment 1 was chosen, the argument was "because you can live more" but without clarifying the relationship with the data. In another class of responses where treatment 2 was chosen, the arguments alluded to stability, regularity or control, for example, "because it is more stable and it can be said that is more effective"

Unistructural level. In the Unistructural level have been classified responses where is taken into account only a relevant datum and the treatment chosen is consistent with the argument. A kind of answers in this level are based on consideration of an extreme value, for example, in Figure 10 a student chooses treatment 2 and argues "because there are likely to live 6.8 years or more", that is, compares the minimum values of each treatment and selects the greatest of them.

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? 2

¿Por qué? porque vives años buenos años y hay mucha probabilidad de que te salga 6.8 para arriba

Figure 10. Unistructural level answer.

In the Unistructural level we have also included responses in which the decision is based on the observation of the modes, and the largest of them is chosen, for example, a student chooses treatment 3 “because 3 patients reach to live 7.1 years with this treatment”. Finally, another type of response included at this level consists of adding the times of each treatment, but with errors on the sum, so that students do not realize of the equality of results. Then they choose the treatment in which data accumulate the greatest amount. For example, in Figure 11 is shown that the student got 71, 70.4 and 70.4 as a result of summing the data of Treatments 1, 2 and 3 respectively. Consequently he chose treatment 1, “Because with this treatment you live a longer time”

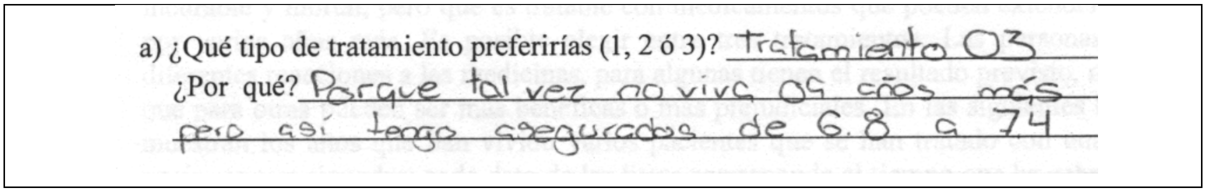
Tiempo en años (Tratamiento 1)	Tiempo en años (Tratamiento 2)	Tiempo en años (Tratamiento 3)
5.2	6.8	6.8
5.6	6.9	6.8
6.5	6.9	6.9
6.5	7.0	7.0
7.0	7.0	7.0
7.0	7.0	7.1
7.0	7.1	7.1
7.8	7.1	7.1
8.7	7.2	7.2
9.1	7.4	7.4
71	70.4	70.4

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? El 1.

¿Por qué? Porque hay más tiempo de año de vida

Figure 11. Unistructural level answer (mistake in the sum).

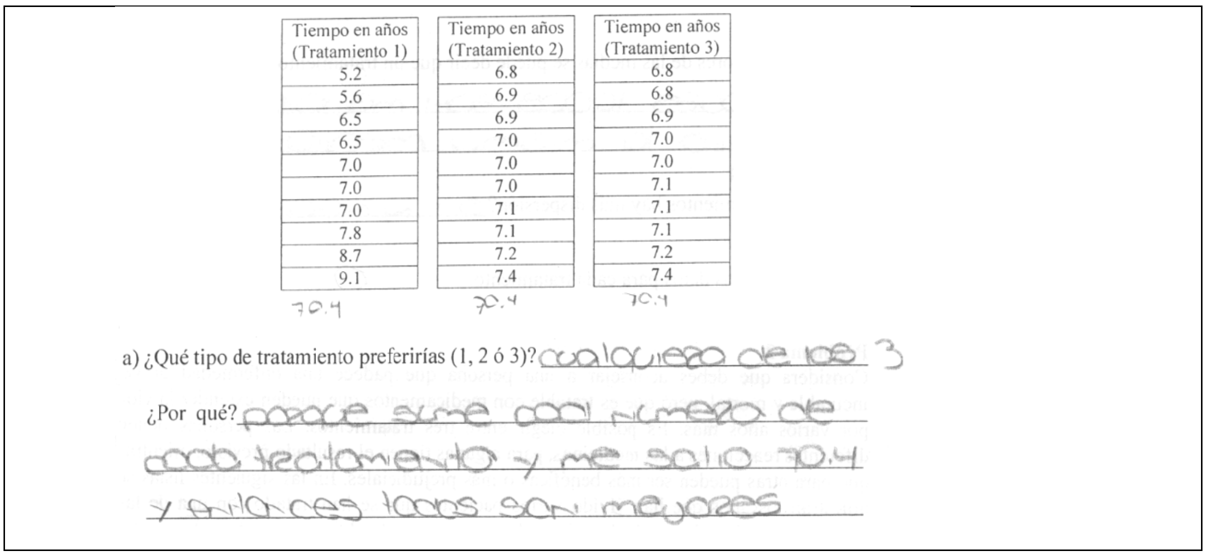
Multistructural level. In the multi-structural level responses were divided into two types. The first type consists of the answers that consider the two extreme values of each dataset. For example, in Figure 12 is shown that the student chose the treatment 3 “Because maybe I will not live nine years but I have secured from 6.8 to 7.4”. Although not mentioned, it appears that the student perceives the risk involved in the first treatment (the possibility to live only 5.2 years) because it gives up the opportunity to live 9, "ensuring" live at least 6.8 years


a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? Tratamiento 03

¿Por qué? Porque tal vez no vivo 09 años más pero asi tengo asegurados de 6.8 a 7.4

Figure 12. Multistructural level answer (considering extreme values of each dataset).

In the second type of responses, the values of each data set are summed and the results are compared, getting to the conclusion that is the same follow any of the three treatments. In Figure 13, the student said that any of the three treatments can be chosen "because I added the data from each treatment and I got 70.4 and then all are better".

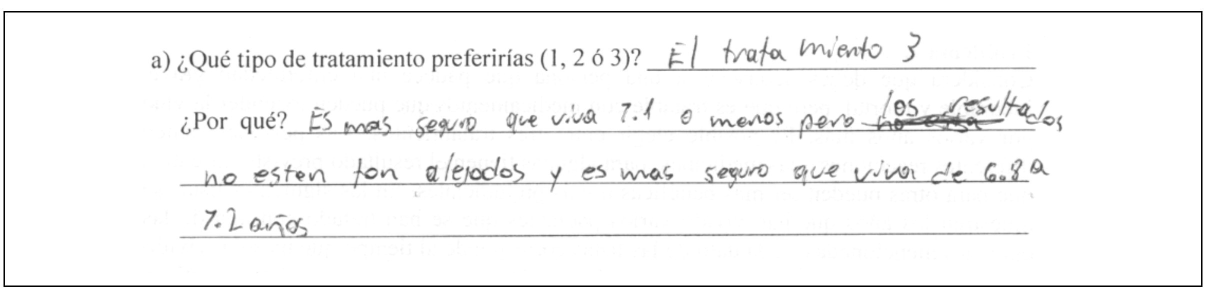

Tiempo en años (Tratamiento 1)	Tiempo en años (Tratamiento 2)	Tiempo en años (Tratamiento 3)
5.2	6.8	6.8
5.6	6.9	6.8
6.5	6.9	6.9
6.5	7.0	7.0
7.0	7.0	7.0
7.0	7.0	7.1
7.0	7.1	7.1
7.8	7.1	7.1
8.7	7.2	7.2
9.1	7.4	7.4
70.4	70.4	70.4

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? cualquiera de los 3

¿Por qué? porque sume cada numero de cada tratamiento y me salio 70.4 y entonces todos son mejores

Figure 13. Multistructural level answer (the values of each data set are summed).

Relational level. In the relational level we include responses that supported its decision taking into account center and dispersion of each dataset. Only two cases were included in this level. In Figure 14, treatment 3 was chosen "because it is more likely to live 7.1 years or less, but the results are not so far apart and are more likely to live from 6.8 to 7.2 years".


a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? El tratamiento 3

¿Por qué? Es mas seguro que viva 7.1 o menos pero ~~no esta~~ los resultados no esten tan alejados y es mas seguro que viva de 6.8 a 7.2 años

Figure 14. Relational level answer.

Level	Pre-test frequency	Post-test frequency
Prestructural	56 (64 %)	47 (61 %)
Unistructural	21 (24 %)	12 (15 %)
Multistructural	10 (11 %)	16 (21 %)
Relational		2 (3%)
Total	87	77

Table 3. Frequencies by level

Conclusions

On the Table 4, we can see that there was in average a decreasing of 15% in Prestructural level from the pre-test to post-test. Several students answered in post-test using data while in pre-test they gave circular arguments. The frequencies of responses classified in Unistructural level in average decreased in 1.34%, although the frequencies to the question 1b increased from 9 to 14. This is because some students improved their perception of dispersion in data, but they were unable to relate this notion with choosing an option in the other problems. An increase in average of 15.67% from pretest to posttest was registered in the frequencies of responses classified at Multistructural level. Most of new responses classified at this level were based on the procedure of seeking the game or treatment where the sum of the data is greater than the sum of the data of the other game; in most of these the conclusion was that the games were equivalent. In the rest of responses in this level students took into account the maximum and the minimum of each set, then they perceived the risk and choosing according to their risk attitudes (risk aversion or risk seeking); in these cases no mention of mean was done. In only two responses to question 2 of the posttest, classified in Relational, students took into account the center (mode) of the sets and observed that one of these involved greater risk; for this, students compared the maxima and minima.

	Question 1a		Question 1b		Question 2a	
Level	Pre-test frequency	Post-test frequency	Pre-test frequency	Post-test frequency	Pre-test frequency	Post-test frequency
Pre-structural	77 (89 %)	54 (70 %)	74 (85 %)	48 (62 %)	56 (64 %)	47 (61 %)
Uni-structural	6 (7 %)	4 (5 %)	9 (11 %)	14 (18 %)	21 (24 %)	12 (15 %)
Multi-structural	3 (3 %)	19 (25 %)	2 (2 %)	13 (17 %)	10 (11 %)	16 (21 %)
Relational	-	-	-	-	-	2 (3%)
No response	1 (1 %)		2 (2 %)	2 (3 %)		
Total	87	77	87	77	87	77

Table 4. Frequencies by level

Our main result is a hierarchy SOLO which describes the patterns of reasoning emerging from the students' efforts to respond to problems. In Table 5 such hierarchy is presented.

Hierarchy SOLO for problems

In general, predominate Prestructural responses; students whose answers are classified at this level understand what is asked them and they make a choice but fail to use the data to support their preferences. However, there are students who see in a single value of each data set (maximum, minimum or mode) a key to make a decision. These responses have been classified in the Unistructural level; they prefigure the valid scheme of solution. The value chosen is one that students consider a representative of the set. In responses of Multistructural level, a step forward towards the solution scheme is given, since more than one value of each data is considered. Two main strategies were identified: 1) compare the sum of values of each set data, 2) take into account the maxima and minima. Each of these strategies is an early or primitive form of the two main statistical tools of the case: the mean and dispersion. The second strategy led some students to perceive the risk. Finally, In Relational level responses, both strategies are used and the decision is made according of attitude towards risk. In these responses a scheme of solution is complete.

Level	Description
Prestructural	One option is chosen but without justification or with a circular response as "because is the better".
Unistructural	The maximum, minimum or mode of each data list is observed and compared
Multistructural	The sums of the data lists, or mean, are compared or both the maximum and minimum are considered and, in this case, risk is perceived.
Relational	The sums or means of the data lists are compared and range (or another measure of dispersion) are considered; the final decision is influenced by preferences about risk

Table 5. SOLO hierarchy

As a final commentary we would like to comment that the indeterminacy of the answers to many of the questions that emerge in Probability and Statistics may be a major cause of the frequent distrust and even rejection toward the discipline by students. In general, an indeterminate response is not considered a satisfactory one; however, good answers that are found in probability and statistics are in some way indeterminate due to the random nature of the phenomena modeled. In the problems that we have reviewed may be not convincing that the choice of a game or treatment is not completely determined by the behavior of the data, but also depends on the solver attitude towards risk. This relativity disturbs those who believe that science must give absolute and conclusive answers to the problems that arise on it. Relativity of the responses may obscure the main point which is that the analysis of the ranges, and more generally the analysis of variation, provides information about the risks involved and therefore helps to make rational decisions. The use in teaching of problems as those treated in this study can help the students to construct schemes for assessing the results of the statistical analysis and help them to retreat from certainty in a profitably way.

References

Bakker, A. & Derry, J. (2011).Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13, 5-26.

Biggs, J. B. and Collis, K. (1982) *Evaluating the Quality of Learning: the SOLO taxonomy*. New York: Academic Press.

Biggs, J. & Collis, K. (1991).Multimodal learning and the quality of intelligent behaviour. In H. Rowe (Ed.), *Intelligence, Reconceptualization and Measurement* (pp.57–76). New Jersey: Laurence Erlbaum Assoc.

Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge MA: Harvard University Press.

Burrill, G., &Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics Challenges for teaching and teacher education:* A joint ICMI/IASE Study (pp. 57-69). New York: Springer.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6, 187-205.

Kahneman, D. &Tversky, A., (2000).Choices, values, and frames. In D. Kahneman& A. Tversky (Eds.), *Choices, Values, and Frames*. Cambridge: Russell Sage Foundation.

Lehrer, R., & Schauble, L.(2007).Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. Lovett & P. Shah (Eds.).*Thinking with data* (pp.149-176). Mahwah, NJ: Lawrence Erlbaum Associates.

Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2&3), 131-156.

Reading C. (2004). "Student Description of Variation While Working with Weather Data", Statistics Education Research Journal, Vol. 3(2), 84-105.

Shaughnessy, J.M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph and K. Carr (Eds.), *People in mathematics education* (Vol. 1, pp. 6-22). Waikato, New Zealand: Mathematics Education Research Group of Australasia.

Watson, J.M., & Moritz, J.B. (2000).Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31,44-70

Watson, J. M., & Kelly, B. A. (2004). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics*, 57, 121-144.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

Wild, D. J., & Pfannkuch, M. (1999).Statistical thinking in empirical enquiry. International Statistical Review, 67(3), 223-265.

Promoting Risk Taking in Mathematics Classrooms: The importance of Creating a Safe Learning Environment

Sashi Sharma

University of Waikato, New Zealand

Abstract: Students beliefs and attitudes towards risk taking can impact on their mathematics learning and performance. However, at present, risk is not established in the field of mathematics education. The challenge for mathematics teachers in developing their students' risk taking dispositions is to choose appropriate activities and tools that match this concept and the learning needs of the students. This paper describes some research-based ideas for promoting risk taking behaviours in a mathematics classroom. It presents interactional pedagogical strategies from a design collaborative research conducted at one secondary school. As part of the learning activities, students critically evaluated statistical investigations undertaken by others. Students had to take risks: ask critical questions, construct statistical arguments and respond to others' arguments in face of criticisms. The findings are interpreted in relation to recent writing about students' abilities to take risks in mathematics classrooms. The final section considers the issues arising out of the paper and offers suggestions for meeting these challenges.

Keywords: mathematics education, risk taking, collaborative research, secondary school students, critical statistical literacy, implications for research.

> *Do the one thing you think you cannot do. Fail at it. Try again. Do better the second time. The only people who never tumble are those who never mount the high wire. This is your moment. Own it.*
>
> ~Oprah Winfrey

Risk abounds in our everyday life and workplace. Risk can delight, annoy and engage us. As the above quote suggests people need to be aware of the importance of risk taking. Notions of risk appear in various messages that we encounter such as when receiving forecasts of medical, financial or environmental risks from media research reports and public officials. In such situations, we have to make decisions in the presence of conflicting goals and constraints (Schlottmann & Wilkening, 2011; Shapira, Nattinger, & McHorney, 2001). In recent times, risk taking and risk-aversion has drawn increased attention of educators and researchers for a range of reasons (Anthony & Walshaw, 2007; Fesser, Martignon, Engel & Kuntze, 2010; Martignon & Kurz-Milcke, 2006; Till, 2014). These include recognition of its importance for solving differences and reaching consensus, as well as its central role in thinking and making informed decisions (Gaissmaier & Gigerenzer, 2008; Galesic & Garcia-Retamero, 2010; Martignon & Krauss, 2009; Schlottmann & Wilkening, 2011). Indeed, to learn and grow people must take risks, but most people will not take risks in an emotionally unsafe environment.

Moreover, there has been significant shifts in the way teaching and learning of mathematics is conceptualised internationally. The American policy document, Principles and Standards for School Mathematics states that if students are to learn to "construct mathematical arguments and respond to others" arguments, then creating an environment that fosters these kinds of activities is essential (National Council of Teachers of Mathematics, 2000, p. 18). This reform teaching aligns with vision promoted by NCTM's Professional Teaching Standards (National Governor's Association for Best Practices and Council of Chief State School Officers. 2010) that encourages

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 290-306

students to communicate mathematical ideas and nurtures intellectual risk-taking by students. However, reasoning at complex cognitive levels and risk taking through mathematical discourse is not something many students are able to achieve easily without adult mediation (Anthony & Walshaw, 2007; Cobb & McClain, 2004; French, 2009; Hunter, 2010).

While there has been ample discussion of students at risk (Clark, 2001; Franco, Sztajn, & Ortigão, 2007; Lubienski, 2007; Winsor, 2007), mathematics education at risk, and nations at risk (Center for the Study of Mathematics Curriculum, 2005; Nasir,, Hand, & Taylor, 2008; Wagner, 2008), risk assessment (Gigerenzer, 2002; Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2009; Martignon & Krauss, 2009), there has been limited discussion of students' willingness to take risks in mathematics classrooms where they are asked to solve open-ended tasks (Lubienski, 2007; Sullivan, Mousley, & Zevenbergen, 2006). Individuals who have different perceptions of risk are more or less likely to do well in these environments (Clifford, 1991; Hills, Stroup, & Wilensky, 2005). It follows that learning environments must take into account individual student's and teachers' risk-taking or risk-averse behaviours. There is very little research on students' mind-sets and motivation for risk tolerance in mathematics education. In 2010, for the first time, risk was included in a session of the ICOTS, as a topic for instruction in school in connection with statistics education (Fesser et al., 2010).

This paper describes some research-based ideas for promoting risk taking behaviours in a mathematics classroom. It presents interactional pedagogical strategies from a design collaborative research (Sharma et al., 2011) conducted at one secondary school. As part of the learning activities, students critically evaluated statistical investigations undertaken by others (Ministry of Education, 2007). Students had to take risks: ask critical questions, construct statistical arguments and respond to others' arguments. Teachers used a range of strategies to initiate, sustain, direct, shift, and conclude mathematical discussions.

The first section reviews literature on risk taking and provides the theoretical framework taken in this paper. Theoretical standpoint of this paper is drawn from a sociocultural perspective which supports a view of mathematical teaching and learning as inherently social and founded on active participation in communicative reasoning processes. The section explains the importance of mathematical communication and questioning in teaching and learning mathematics. The second section will draw upon Sharma et al (2011) study to explain pedagogical strategies used in the study to promote risk taking. Episodes are provided to illustrate the demand for the teacher to engage in a number of pedagogical actions in order to maintain student participation and engagement in the discourse community. The final section outlines some implications and offers suggestions for teaching and further inquiry into risk literacy.

Literature Review

Research on Risk Taking

Understanding how people think about risk and their attitudes towards risk taking is important for educators and policy makers (Xie, Wang & Xu, 2003). It is widely assumed that people differ considerably in their motivation and attitudes towards risks, ranging from cautiousness to risk-seeking and even pleasure in risk-taking (Burrill , 1998; Rohrmann, 2005). However, there is no convincing evidence from these studies that this is a general trait (Hills, Stroup, & Wilensky, 2005; Rohrmann, 2005). According to Rohrmann, risk attitudes are multi-dimensional and that risk orientations are not consistent across domains and the motivation for accepting risks depend on the context. There is very little research on students' mind-sets towards risk taking in mathematics education. To address this issue, Atkins, Leder, O'Halloran, Pollard and Taylor (1991) investigated the tendency of students to take risks in mathematics examination (in multi-choice format). Atkins, et al. (1991, p. 297) defined risk taking as "the preparedness of a student to attempt to answer a question when not certain of the result".

Specifically, data from the Australian Mathematics Competition were analysed to measure risk taking by groups of secondary students by gender, school year and achievement level. An example of a problem from their study is given below.

Mary wanted to divide a certain number by 4 to get an answer. However, she used the calculator incorrectly and multiplied by 4 instead and got 60. The correct answer would be

(A) 3.75 (B) 15 (C) 4 (D) 12 (E) 240

The statistics for this question were as follows:

(A) 34% (B) 42% (C) 5% (D) 7% (E) 4% No response 6%

Atkins et al. (1991) explained that the incorrect responses ©, (D) and (E) were more likely to be due to risk taking techniques whereas the high response for the alternative (B) was probably due to participants who believed that they had answered correctly but had not read the question carefully.

The examination paper consisted of 30 multi-choice questions. The first 10 questions were worth 3 points, the next 10 were worth 4 points and the last 10 were worth 5 points. The range of possible scores was from 0 (for 30 incorrect responses) to 150 (for 30 correct responses). The authors measured risk by finding the average value of penalty marks which ranged from 0 to 30 for an individual student.

The measures showed relatively high values for the lower level grades decreasing each year to grade 12. Males consistently obtained higher z values than females except for grade 10. The researchers concluded that among high school students, the highest risk taking behaviour takes place among year 7 students and steadily decrease until year 12. Findings regarding gender differences resonate with the findings of Fesser et al. (2010). Data from Fesser et al. sample exhibited interesting gender differences, girls tended to be more risk averse than boys in a typical gamble and a ludo game situation. However, Fesser et al. also reported that younger students lacked effective tools for assessing and communicating risk which is inconsistent with the findings of Atkins et al. Fesser et al. findings motivate the development of research programmes for implementing risk assessment and risk communication as a mathematics topic in school curricula. Moreover, Fesser et al. argued that an early familiarisation with risk as a topic in school may strengthen intuitions on risk and convey competencies for sound decision making in risky situations.

It appears that children can be taught to be more risk seeking in their activities. Dehaene, Izard, Spelke and Pica (2008) contributed to this picture through their examination of whether the move from risk aversion (logarithmic) to risk taking (linear mapping) was a result of formal schooling or a natural process of brain maturation. To study this, they undertook some number-mapping exercises with the Mundurucu, an Amazonian group with little access to education or other instruments that may affect their perception of numbers (such as maps and rulers). The researchers concluded that the education of children in mathematics tended to increase their risk tolerance through changing the way they see numbers. The conclusions conform the findings of Fesser et al. (2010).

Streimatter (1997) claims that in order for people to understand who they are, they must experiment and take risks. She further adds that one appropriate place for young people to take risks as they explore who they are is in the mathematics classroom. Streimatter argues that "without taking academic risks, asking or answering questions in the classroom a large part of students' lives may be excluded from their conscious or subconscious deliberations during this period of identity" (p. 18). Indeed, students who are active participants in their own education tend to be higher achievers. It must be noted that although asking and answering questions in class may seem rather trivial compared to risk taking behaviours, for some students, these behaviours involve substantial risk tolerance.

Streimatter (1997) explored the dynamics of a girls-only class in a public co-educational middle school. Her classroom observations and interviews with girls revealed that girls were more likely to ask and answer questions about subject matter in the mathematics class than they were in their other classes which were coeducational. The girls also said that the girls-only setting enhanced their ability to learn mathematics and their views of themselves as mathematicians.

Indeed in numerous areas, including financial decisions, health/safety, and recreation females are found to be more risk-averse than males (Weber et al., 2002). These gender differences in everyday decisions may also carry over to educational performance in mathematics. For example, a meta-analysis performed by Hyde et al., (1990) (cited in Hills, Stroup, & Wilensky, 2005), assessing the performance of over 3,000,000 students, suggests that females outperform males in most grades, although, this gender gap has been shrinking. However, they also noted a significantly lower performance of females in mathematical problem solving at the high school level.

From the above discussion, it appears that gender influences students' risk seeking behaviour. If gender issues are prevalent in risk taking behaviour in-inquiry based and student-centered opportunities may be inappropriately biased towards a specific gender. As well, if there is a variation in risk-seeking or risk-averse behaviour across students, then proposals to introduce risk-intensive curriculum must be appropriately guided to meet students at appropriate comfort levels.

According to a number of writers (Ames &Archer, 1988; Burrill, 1998); Elliot, 1991; Hills, Stroup, & Wilensky, 2005; Meyer & Turner, 2002; Stipek, Salmon, Givvin, Kazemi, Saxe & MacGyvers, 1998) tolerance for risk is directly related to students' perceptions about their own goals (performance or mastery). Elliot (1999) states that performance goals focus on the demonstration of competence relative to others, whereas mastery goals focus on the development of competence. Elliot adds that performance goals are hypothesised to be linked to a negative set of processes and outcomes, for example, withdrawal of effort in the face of failure, decreased task enjoyment, surface processing of study materials. On the other hand, mastery goals are hypothesised to be linked to a positive set of processes and outcomes, for example, persistence in the face of failure, enhanced task enjoyment and deep processing of study materials.

Ames and Archer (1988) claimed that individuals who see learning as the ultimate goal and who are less worried about risks of assessment are more likely to pursue challenging material and thereby engage the risk and confusion associated with learning. However, when success is normatively defined, then both self worth and the perception of the material are at stake. Hence, in a given activity, students and teachers may perceive performance or mastery goals differently, and this may affect their willingness to participate or engage in 'risky' learning situations.

The above claims are consistent with the findings of Dweck (2000) and Burrill (1998) who state that how teachers approach a curriculum may be a consequence of their perception of associated risks. Dweck explained that students who are more likely to avoid risk taking have a performance orientation, meaning they seek social affirmation rather than understanding of the content. Dweck claims that teachers may inadvertently encourage such responses by affirming easy successes and by failing to affirm effort. According to the Burrill's (1998) review of the Third International Mathematics and Science Study, 61% of lesson goals among United States teachers focused on skills, 22% focused on thinking, and 6% were on test preparation. Burrill adds that though many aspects are undoubtedly involved in these decisions, the above evidence would suggest that students and teachers perceptions of risk associated with performance and the complexities associated with thinking are likely to be significant factors. If students are unwilling to tackle challenging content with its associated risks of confusion and failure because of external performance evaluations, then these performance evaluations inhibit learning advocated by current reforms (NCTM, 2000). With or without these performance evaluations, students and teachers are likely to show natural variation in their willingness in challenging situations

Hills, Stroup and Wilensky (2005) explain that student-centered learning implies perceived risks by both teachers and students. By nature open-ended problems have multiple entry and exit points, require high levels of cognitive demand and may have fuzzy criteria for correctness (Ferguson, 2009; Sullivan & Mornane, 2014). It follows that people who have different perceptions of risk (Hills, Stroup & Wilensky, 2005; Rohrmann, 2005) are more or less likely to do well in these environments. Appropriate educational environments must therefore take into account students and teachers risk-taking or risk-averse attitudes.

Using a risk-eliciting investment game, a survey of preferences, and a data-blind evaluation of participation in class discussion, Hills, Stroup, and Wilensky (2005) investigated the hypothesis that there are risk-seeking and risk-averse pre-service teachers and that this risk is conserved across activities. They found strong correlations between risky behavior, risk preference, and willingness to engage in both in class discussions and open-ended problems in general in mathematics and science. Approximately 65% of pre-service teachers only engaged in discussions when directly asked a question and these teachers also strongly preferred non-constructivist type of mathematics and engaged in 'safe' behavior in the investment game. Risk-seeking students in the game were more likely to engage in discussion and preferred open-ended and confusing activities. Unlike Streimatter (1997) and Atkins et al. (1991) studies, gender was not associated with students' risk preference except males were more prone to seek open-ended science activities.

Meyer and Turner (2002) investigated academic risk taking behaviour of students in an upper elementary mathematics classroom during several project-based activities. The researchers used two surveys, one that assessed academic risk taking and the extent of students response to failure and the other assessed individual goals, self-efficacy and strategy use. In addition, they videotaped classroom lessons and interviewed students before and after the learning goals In contrast, the risk avoiders note more negative emotions, limited social support and very different goals and strategies for their project. project to understand how the students' beliefs were related to students' actions. They found that the risk takers approached the project's learning opportunities with positive feelings and in ways that met their goals.

Reforms in mathematics education

Over the last two decades, mathematics education reform has called for a shift in approach to the teaching and learning of mathematics in order to meet the needs of a knowledge society. A major aspect of this reform is the move from traditional teacher-centred classrooms focusing on content, to classrooms where classroom discourse and conceptual development are facilitated (Cobb, Boufi, McClain, & Whitenack, 1997; Fraivillig, Murphy & Fuson, 1999; Franke, Kazemi & Battey, 2007; Hunter, 2010; Mercer & Sams, 2006; Silver & Smith, 1996, Stein, 2001). The reforms have recommended that teachers emphasise focus on processes and seeking solutions rather than following a set of solution path (Begg, 2009; French, 2009; Hunter & Anthony, 2011; Stein, 2001). The reforms suggest giving students opportunities to engage in mathematical conversations, incorporating students inadequate solutions into teaching and giving feedback rather than grades on assignments. Dalton (1990) adds that risk taking involves guessing, sharing ideas with others that might involve criticisms or failure or defending ideas. Teaching approaches that foster these learning goals also engender a willingness to engage in challenging tasks and to take risks (Begg, 2009; Sullivan & Mornane, 2014).

The above recommendations resonate with principles of socio-cultural theories combined with elements of constructivist theory which provide a useful model of how students learn mathematics. Constructivist theory in its various forms, is based on a generally agreed principle that learners actively construct ways of knowing as they strive to reconcile present experiences with already existing knowledge (Confrey & Kazak, 2006; Lerman, 2006; von Glasersfeld, 1993). Students are no longer viewed as passive absorbers of mathematical knowledge conveyed by adults; rather they are considered to construct their own meanings actively by reformulating the new information or restructuring their prior knowledge through reflection (Cobb, 1994).

Another notion of socio-cultural theory derives its origins from the work of socio-cultural theorists such as Vygotsky (1978) and Lave and Wenger (1991) who suggest that learning should be thought of more as the product of a social process and less as an individual activity. There is strong emphasis on social interactions, language, experience, collaborative learning environments, catering for cultural diversity, and contexts for learning in the learning process rather than cognitive ability only. Children learn through social interaction, by talking, explaining, listening, and actively exploring concepts with their peers in whole-class and small-group situations. Through the process of verbalising, including asking/responding to questions, children learn to make connections between concrete and abstract thought (Perso, 2003). Hence, the development of mathematical understanding requires learning contexts in which children can discuss and reflect on their construction of knowledge (Cobb et al, 1997; Franke et al, 2007; Hunter & Anthony, 2011; Mercer & Sams, 2006; McFeetors & Mason, 2006).

The sociocultural perspective has led to an increased attention to mathematical communication and questioning in curriculum documents and materials produced for teachers (Anthony & Walshaw, 2007; Lampert & Cobb, 2003, Lerman, 2006; Ministry of Education, 2007; National Council of Teachers of Mathematics [NCTM], 2000). However, this can only happen if teachers create social norms in their classrooms that give students the confidence to ask questions, discuss with others and to listen and respond actively to the ideas of others (Cobb & McClain, 2004; Franke et al., 2007; Hunter, 2010).

Studies on reform teaching offer contradictory results. A study in England by Jo Boaler (2008) used a fixed average (low income) student socioeconomic status (SES) in two schools while examining varied teaching approaches. When Boaler compared students' performance between schools, her results pointed to reform teaching as beneficial to low SES students. In a study in the United States by Lubienski (2007) all the students learned math by a reform approach but the SES of the students varied. Lubienski's analysis of data from students of differing SES in the same classroom showed that reform was problematic for the low SES students. Lubienski suggested that these differences between the teacher's intentions and students' perceptions were manifestation of cultural confusion

Stipek, Salmon, Givvin, Kazemi, Saxe and MacGyvers (1998) claim that students willingness to take risks can be seen in students reactions to having difficulty in regular classroom contexts. Students may not seek help because of the risk of being perceived as being dumb in asking questions, they may give up, persist on ineffective strategies instead of seeking assistance. Indeed, without explicit discussion of the structure of the discourse-how it works, its norms and rules some students may not be able to take risks and participate fully in the rich conversations.

Recent research has shown that intervention can assist low SES students to learn about the value of classroom mathematical discourse. Hunter (2010) investigated discourse patterns within two year 5 classes of mostly Pasifika and Maori children attending a low SES school. The teachers used explicit strategies to develop and maintain student participation in the discourse community because these were not familiar to the children. Hunter (2010) argues that children from minority groups need to be encouraged by their teachers to participate in mathematics discussion and taught how to do this.

Sharma et al (2011) study

The following inter-related research questions guided the study:

1. How can we develop a classroom culture where students learn to make and support statistical arguments based on data in response to a question of interest to them?

2. What learning activities and tools can be used in the classroom to develop students' statistical critical thinking skills?

A design research approach (Cobb, 2000) with its progressive cycles of testing and revision was used for this study. Design research generally involves cycles of three phases: a preparation and a design phase, a teaching experiment phase, and a retrospective analysis phase. Teachers were key stakeholders in this collaborative research project consistent with Kieran, Krainer and Shaughnessy (2013).

Preparation for the teaching experiment

This phase consisted of literature review and the first attempt at reformulating a Hypothetical Learning Trajectory (HTL). Then, the research team proposed a sequence of ideas, skills, knowledge and attitudes that they hoped students would construct as they participated in activities. The team planned activities to help move students along a path towards the desired learning goals. As part of the activities, students evaluated statistical investigations or activities undertaken by others including data collection methods, choice of measures and validity of findings (Ministry of Education, 2007).

Teaching Experiment

The teaching tool place in regular classrooms and as part of mathematics teaching. The teaching activities were spread over up to two weeks to suit the school schedule. The research team was involved in designing, teaching, observing and evaluating sequences of activities. There were two cycles of teaching experiments. Their goal was to improve the design by checking and revising conjectures about the trajectory of learning for both the classroom community and the individual students.

Data Collection

During the teaching experiment, the data set consisted of video-recordings of classroom sessions conducted during the research, copies of all the students' written work, audio recorded mini-interviews conducted with students, and set of field notes of the classroom sessions. Semi-structured interviews were also conducted with a selected number of students from each class while the design experiment was in progress. These interviews were scheduled after class sessions and focus on students' interpretation of classroom events with a particular emphasis on the identities they were developing as consumers of statistics. Each teacher-researcher kept a logbook of specific events that took place during the data collection period. The team was engaged in conscious reflection and evaluation of situations as they unfolded.

Data Analysis

The research team performed a retrospective analysis after each lesson to reflect on and redirect the learning trajectory. In addition the team performed analysis of the HTL after an entire teaching experiment has been completed. During this phase the team developed specific instructional theory to help future instruction. The continually changing knowledge of the research team created continual change in the hypothetical learning sequence.

Results and Discussion

A number of learning activities were trialled in our study. In this section, music survey activity which intended to capitalise on students' interest in music, is used to discuss the results.

Activity – Music Survey

The purpose of this activity is to introduce students to some critical questions that will help them evaluate surveys undertaken by otheres. The activity provides opportunities to students to express their opinions and concerns both verbally and in writing.

Start by introducing the context of the activity by asking questions such as:

How many of you like music? What type of music do you like the most? How much do you spend on music?

Then present the scenario below. Students can read the music survey scenario individually and make a note of any difficult words. These words can be posted on the whiteboard. Students could be encouraged in pairs to represent the words through interpretative drawings and labels.

Survey shows Recorded Music Appeals to Teens

A marketing research company was asked to determine how much money teenagers (age 13 – 19) spend on recorded music (cassette tapes). The company randomly selected 80 malls located around the country. A researcher stood in a central location in the mall and asked passers-by who appeared to be the appropriate age to fill out a questionnaire. A total of 2 050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends $1155 each year on recorded music.

Ask students to answer the questions below. Having students working in small groups enables discussion of their opinions.

1. Who carried out the survey?

2. What was the purpose of the survey?

3. How was the survey done? Why do you think this method was used?

4. What was the target population?

5. What are the main findings of the survey?

6. How do you think the average was calculated? Is the average a good estimate of the spending of all teenagers? Why?

7. Does the data support the headline? Justify your answer.

8. Comment on two concerns you have about the survey.

Encourage students to write down their response for each question. You may want to have groups present their decisions to the class.

When we used the music survey activity, concerns raised by students about the sample selection process used in the study were:

> *They only asked passers-by in the mall. As Mr T says, this might bias the results because teenagers outside the mall were not asked.*
>
> *How did they calculate the average? Did they get rid of odd numbers [outliers]? Some teenagers spend more money than others.*
>
> *They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.*

The last student quote above reinforces to us that students can struggle with thinking of the sample size in relation to the size of the country, rather than in relation to the representativeness of the sample. However, for the above scenario, theirs is a valid concern because hardly any information had been given regarding the target population.

Like the activity above, most of our classroom activities included group and whole class discussion of the data. Typically, we used a small group activity (2-4 members) in which the students worked on tasks together and then reported back to the whole class. The groups were organised to include students with a range of mathematical, statistical, and language abilities because we believed that varying insights would enhance overall understandings.

During group work learning opportunities arose for students from collaborative dialogue and resulting of differing points of view. Working in groups also provided less confident or less able students with opportunities to explain , question, agree and disagree and test their thinking in a less threatening context before engaging in class discussions. Our findings are consistent with the views of Hunter (2010) and Begg (2009) who illustrated the importance of small group discussion as a means for students to rehearse their explanations, justifications and analysis of their solution strategies as the students prepared for questioning and challenge from the whole class.

The teachers ensured that students understood and adhered to effective group problem solving practices, including listening, writing, answering, questioning, and critically evaluating information. The teacher in the following transcript explains their expectations:

> *When working in the group, first of all each one of you has to say what your concerns are. Then I want you to come up with a group response. You should not only say bias, headline and questions, you need to clearly explain your reasons. Remember - at the end of your group discussion, everyone in the group needs to be able to explain why you chose that response. Also, anyone in the group can be asked to present to the whole class.*

During reporting back the whole group was required to stand with the reporter and share the responsibility for explaining their reasoning and responding to any questions from the class. These norms encouraged the students to work together, communicate, and be responsible for the learning of everyone in the group. The teachers took time to remind students of effective group practices (e.g., ways of agreeing and disagreeing and how to present to the class):

> *The group is not finished until everyone in the group can explain and defend their answer. When we come together as a class to share our ideas, I will simply ask any one student to report on why they agree or disagree with a particular statement.*

Anthony and Walshaw's (2009) research provides evidence that learners enjoy learning in communities of mathematical enquiry as they gain a variety of strategies that can be used to solve problems and they feel their belonging to a group that allows them to engage with their own sense-making and the sense-making of others. Our study shows that working in groups provides learners with more than just opportunities to engage with mathematics, it also provides learners with opportunities to develop other competencies identified in the New Zealand curriculum document (Ministry of Education, 2007). For example, students become aware that working with others will support their learning and understanding and the community provides a safe environment where they can take risks asking questions and defending ideas.

Students in our classes were of different language and statistical abilities and activities were designed to ensure students were interacting with each other in order to improve their statistical communication. Support was provided for both reading and writing in statistics. Supports included assisting with vocabulary acquisition, such as using pre-reading and further reading strategies such as shared reading and scanning techniques. Writing support included the use of writing frames, cloze activities and composing responses individually and in groups. The student-student interactions presented various demands on students' literacy skills, as indicated in the following student quote:

> *Because usually, like in normal maths, we don't use literacy … like we use addition, subtraction but we actually need some kind of literacy for the things we do in statistics.*

Questioning was one of the key teaching strategies used in our study. We wanted both the students and the teachers posing questions that would support student learning. The teachers posed prompting and probing questions that diagnosed and extended student thinking - questions that elicited student ideas and encouraged them to explain and justify their contributions in respectful ways (Cobb & McClain, 2004; Franke, Kazemi & Battey, 2007). Students needed to formulate and pose critical questions in ways that assisted them to evaluate statistical statements and reports.

Our students came up with critical statistical questions, as reflected in the following:

The simplest question I want to ask is how they got the information. How were the teenagers chosen? Now that we have talked about random samples in statistics … did they use random sample?

I want to know how many boys and how many girls took part in the research. To show comparison, they need same number of boys and girls.

It seems that among the students interviewed classroom culture was an important factor in how they responded to asking questions. Some students commented that because the teacher explained that students could ask different questions they were more prepared to ask questions as they felt less pressure in terms of making mistakes in front of their peers. Some students even indicated that they were willing to risk making mistakes in front of their peers.

When asked what helped them to try harder in maths each interviewed student reported that the teacher was the most influential factor. One student explained why this relationship was so critical to their learning

I have a good relationship with my teacher, I am not scared to ask him questions. I don't ask other teachers because I am afraid they might get angry

Another theme that emerged was that understanding what was being asked was important when beginning the lesson. Students specifically pointed out that at the initial stage key terms were explained and connections made to their prior knowledge. This an important insight to incorporate into the recommended pedagogies.

The classroom discourse was important for statistical literacy. Most of our classroom activities included group and whole class discussion of the data. This typically involved a small group activity in which the students worked on problems together and then reported back to the whole class. The two teachers took time to remind the students how to work in groups (e.g. how to agree and disagree and how to present to the class). Our results show that students can be taught how to question and challenge in respectful ways as part of classroom discourse.

Students found group work useful:

When you are working alone you just get one point of view and when you are working in a group you get different perspectives of other ideas … how other people are thinking, learning in class

Oh … just because when we work alone we might get it right, we might get it wrong, but if we work in a group we'll get more ideas. We will be able to discuss it with the group.

Context is an important component of statistical literacy. Our findings show that students need exposure to both familiar and unfamiliar contexts. Engagement with context helps students develop higher order thinking skills. However, our results show that some contextual knowledge may be a barrier for some students. Teachers were able to address this in two ways. The first was to start from familiar contexts before moving to unfamiliar contexts. The other was to use contexts of interest to the students.

Our study shows that students concerns about revealing ignorance by asking questions might be overcome in a classroom context that focuses attention on learning rather than just getting right answers. The findings align with the conclusions of Burrill (1998) and Elliott(1999). According to Elliott, learning goals focus on the development of competence or task mastery whereas performance goals focus on the demonstration of competence relative to others. In our study a context, teachers encouraged students to ask questions about media reports, they were concerned with processes such as persistence in the face of failure, deep processing of tasks, risk taking. In additions, classmates were prohibited from putting down a student who was having difficulty posing an appropriate statistical question. The teachers were aware that students who have been shamed by the teacher or another student will not will not take risk and engage in challenging tasks. Students in the class commented:

I not bothered about being wrong. Even if you have something wrong he will listen to you and ask you to pose a better question.

He always reminds us not to put down our classmates.

Sometimes I don't have the right question and answer. It is okay to make mistakes. Mr ... says we learn from our mistakes.

Overall the interviewed students were willing and some cases eager to evaluate statistical tasks. They were aware of the subtleties in the classroom that affected their learning in mathematics and the factors that either assisted or constrained their ability to take risks.

Implications for Teaching and Research

We find that this work is still in its infancy and there are numerous limitations. As the research questions suggest, the original study was focused on developing statistical literacy. The study was not designed to give us insight into risk taking. Thus, results from these studies are tentative. However, these limitations provide an opportunity for future research. The work that has been done provides a road map for future studies.

We envision statistical literacy going beyond calculations. It is more than the ability to do calculations and read tables and graphs. Students should be able to interpret and critically evaluate statistical information and data related arguments. Additionally, they should be able to take risks and communicate their understanding and opinions to others. This has potential consequences in how the teaching of statistical literacy might be altered for greater effectiveness. For example, ample class time should be spent on discussion and reflection rather than presentation of information.

We believe that the nature of the learning environment and classroom culture are major contributors to success for students, and teachers need to put a high priority on building a classroom climate that positively engages all students. Students need to understand the importance of sharing their opinions in order to advance their statistical ideas. It would be valuable for teachers to help students reflect on the purposes of explaining and justifying their thinking to others This is consistent with the latest New Zealand curriculum document that promotes the ideals of having confident, critical and active learners of mathematics (Ministry of Education, 2007).

The ability to interpret and critically evaluate reports that contain statistical elements is paramount in our information laden society. Teachers need to give students some basic foundations for critiquing and evaluating statistically based information that they encounter in daily life. We assume that students can be taught these reasoning skills through using media articles as a springboard into learning about how to evaluate these reports. Consequently they will become familiar with a list of worry questions and apply them to real life examples without prompting, consistent with Gal (2004).

It was clear that specific actions by the teachers to encourage and affirm risk taking along with information about the processes of explaining their thinking in writing and verbally to other class members and the importance of active listening was helpful. Indeed as students move to statistics study at more senior levels an orientation to risk taking and a growth mindset will be useful asset to them.

Teachers have to be prepared in the initial stages for students to be negative towards challenging tasks and be prepared to scaffold their learning in a safe environment in order to become competent mathematical practitioners. Like Anthony and Walshaw (2009) we found that students find it difficult to provide mathematical explanations, to figure out what questions to ask and often lack the confidence to speak in large groups.

Views about statistics teaching and learning have shifted considerably in New Zealand and internationally over recent decades, and it is important for teachers to be kept informed about changes in the ways that mathematical and statistical processes and thinking are being emphasised (Anthony & Walshaw, 2007). It would be useful if schools were to make a point of highlighting the importance of risk taking in mathematics and other learning areas in their interactions with teachers and families.

Not only does this help with mathematics but learning the process of critical thinking and risk taking will expand to other curriculum subjects and wider society. Our study was a long term teaching experiment involving a team of teachers and researchers and it only begins to suggest ways to promote risk taking. More research is needed on the effectiveness of particular sequences of activities, the use of different technological tools and the teachers role in shifting the students towards desired behaviour. Other forms of classroom based research can be used to study and attempt to answer questions about optimal ways of promoting risk taking in mathematics teaching. Studies that give insight into the challenges faced by teachers and learners in inquiry-based classroom become important. Teachers can reflect on and develop their practices. It would seem that ongoing research is valuable to inform teacher practice.

The results have implications for teacher education. Many discussions around presentation of mathematical concepts to students are focussed on avoiding the complexities. Yet it may be that coming to grips with the complexities is exactly what students need. It seems important that teacher educators include at least some consideration of risk taking in their presentations to prospective and practising teachers.

Ideally, it would have been useful to make links between students' classroom experiences with what happened outside of the mathematics classrooms. This could have been achieved by doing observations in other classrooms and by conversing with other teachers. This could have enabled the research team to gain insight into whether the students were transferring the critical skills from the realms of statistics to other learning areas as advocated by Watson (2006).

Comprehensive theoretical work that articulates how motivation and cognition interact within mathematics classroom contexts is needed if understanding risk taking is to move forward. Although our research has borrowed ideas from socio-cultural learning theory that appeared to support our research questions, this framework provided only some of the pieces to the theoretical puzzle that we are trying to investigate here. We need theories that will help us better understand how risk taking is intertwined with motivation and cognition within the context of classroom learning.

Research discussed in this paper indicates that females are found to be more risk-averse than males (Hills, Stroup, & Wilensky, 2005; Weber et al., 2002). These gender differences in everyday and educational settings may also carry over to performance in mathematics classrooms. However, gender gaps in students willingness to take risks associated with higher-level problem solving has not yet been investigated. Additionally, we failed to ask how teachers and students defined risk and how they felt about risk taking in mathematics learning. Such issues could be addressed in future research.

The participants in my study were a fairly small non-random sample from one school. Thus, the findings, in particular the number of students who took risks may or may not generalize to the population of secondary school students as a whole in New Zealand. There is a need for more research with larger, more random samples with different backgrounds to determine how common these behaviours are in the general population.

Furthermore, as mentioned earlier, the findings reported in this paper were part of a larger study (Sharma et al. 2011) which focused on developing critical statistical literacy. Since there had been few studies that focused on risk taking internationally, it was not clear when this study was conducted that the questions discussed in this paper would be as rich and interesting as they were. Now that the risk dimension described in this paper has been identified as possible areas of concern, there is a need for more qualitative and quantitative research focused on a deeper understanding of students' thinking about risk and their risk taking behaviours.

Concluding Thoughts

It is hoped that the findings reported in this paper will generate more interest in research with respect to risk taking, risk perception and risk communication in mathematics education. It will be interesting to explore gender and cultural differences that may impact on students' risk taking behaviour. There is also a need to focus on documenting the challenges and difficulties that researchers face in the process of conducting international studies and how cultural factors can influence researcher activities and research results. Indeed we need to look for new ideas and develop more collaborative and cross cultural research between practitioners and researchers in the future if we are to improve outcomes for all our students. Research methods developed in one cultural setting may not be appropriate in another cultural context (Cao, Forgasz, & Bishop, 2007). Teachers, curriculum developers and researchers need to work together to find better ways to help all students take risks in mathematics classrooms.

References

Ames, C., & Archer, J. (1988) Achievement goals in the classroom: student's learning strategies and motivation processes. *Journal of Educational Psychology*, 80, 260-267.

Anthony, G. & Walshaw, M. (2007). *Effective pedagogy in mathematics/pängarau: Best evidence synthesis iteration [BES]*. Wellington: Ministry of Education.

Anthony, G., & Walshaw, M. (2009). Characteristics of effective teaching of mathematics: a view from the west. *Journal of Mathematics Education, 2*(2), 147-164.

Atkins, W. J., Leder, G. C., O'Halloran, P. J., Pollard, G. H. & Taylor, P. (1991). Measuring Risk taking. *Educational Studies in Mathematics,* 32, 297 -308.

Begg, A. (2009). Learning tasks: More than content, In R. Averill & R. Harvey (Eds.), *Teaching Secondary School Mathematics and Statistics: Evidence-based. Practice* (pp. 11-21, Volume One). Wellington: NZCER Press.

Blais. A. & Weber, E. U. (2008). A Domain-Specific Risk-Taking (DOSPERT) scale for adult population *Judgment and Decision Making*, 1(1), 33–47.

Boaler, J. (2008). *What's math got to do with it? Helping children learn to love their least favourite subject – and why it's important for America*. New York, NY: Viking.

Brown, C. L., Cady, J. A., & Taylor, P. M. (2009). Problem solving and the English language learner. *Mathematics Teaching in the Middle School, 14*(9), 532-539.

Burril, G. (1998). Changes in your classroom: From the past to the present to the future. *The Mathematics Teacher,* 9199), 800-806.

Cao, Z., Forgasz, H. & Bishop, A. (2007). Doing surveys in different cultures: Difficulties and differences-A case from China and Australia. In B. Atweh, A. C. Barton, M. C. Borba, N. Gough, C. Keitel, C. Vistro-Yu & R. Vithal (Eds.), *Internationalisation and Globalisation in Mathematics and Science Education* (pp.301-319). Dordrecht, The Netherlands: Springer.

Center for the Study of Mathematics Curriculum (2005). *A nation at risk: The imperative for educational reform,* National Commission on excellence in education.

Clark, M. (2001). Cross-cultural issues with students from the South Pacific. *Australian Mathematics Teacher*, *57* (1), 17-20.

Clifford, M. M. (1991). Risk taking: Theoretical, Empirical, and Educational Considerations. *Educational Psychologist,* 26(3/4), 261-297.

Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science* (pp. 307–333). Mahwah, NJ: Lawrence Erlbaum.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23 (7), 13-20.

Cobb, P. & McClain, K. (2004). Principles of instructional design for supporting the Development of students' statistical reasoning. In D. Ben-Zvi & J. B.Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 375-395). Dordrecht, the Netherlands: Kluwer.

Cobb, P., Boufi, A, McClain, K & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education, 28* (3), 258-277.

Confrey, J & Kazak, S. (2006). A thirty-year reflection on constructivism in mathematics education in PME. In A. Gutiérrez & P. Boero (Eds.), *Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future* (pp.305-345). Rotterdam: Sense Publishers.

Dalton, J. (1990). *Adventures in thinking: Creative thinking and cooperative talk in small groups*. Melbourne: Thomas Nelson.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures *Science, 320* (5880), 1217-1220 DOI: 10.1126/science.1156540

Dweck, C. S. (2000). Self-theories: *Their role in motivation, personality, and development*. Philadelphia, Psychology Press.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 43(3), 169-189.

Ferguson, S. (2009). New voices. Same task, Different paths: Catering for student 117 diversity in the mathematics classroom. *Australian Primary Mathematics Classroom,14*(2), 32-98.

Fesser, S., Martignon, L. Engel, J., & Kuntze, S. (2010). Risk perception and risk communication of school students: First empirical results from Riko-Stat In C. Reading (Ed.) *Proceedings of the 8th International Conference on the Teaching of Statistics,* Ljubljana, Solvenia: International Statistical Institute and International Association for Statistical Education. Available from: Http://Www.Stat.Auckland.Ac.Nz/~Iase/Publications

Franco, C., Sztajn, P., & Ortigão, M. I. R. (2007). Does reform teaching raise all students' math scores? An analysis of the role that the socioeconomic status of students plays in closing the achievement gap. *Journal for Research in Mathematics Education* 2007, 38(4), 393-419.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225-256). Charlotte, NC: Information Age Publishing.

French, S. (2009). Promoting thinking in the mathematics classroom. In R. Averill & R. Harvey (Eds), *Teaching Secondary School Mathematics and Statistics: Evidence-based. Practice* (pp. 11-25, Volume Two). Wellington: NZCER Press.

Gaissmaier, W. & Gigerenzer, G. (2008). Statistical illiteracy undermines informed Shared decision making. *ZEFQ,* 102, 411-413.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In J. B. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking,* (pp. 47-78). Dordrecht, The Netherlands: Kluwer.

Galesic, M., & Garcia-Retamero, R. (2010). *Statistical numeracy for health. A cross-culturalcomparison with probabilistic national samples.* Archives of Internal Medicine, 170(5), 462-xxx http://archinte.jamanetwork.com/article.aspx?articleid=415709#RESULTS

Garfield, J. B. & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics*, *31*(30), 72-77.

Gigerenzer, G. (2002) *Calculated risks: how to know when numbers deceive you*. New York: Simon and Schuster.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. W. (2009). Knowing your chances. *Scientific American Mind*, *April/May*, 44-51.

Gigerenzer, G., & Gray, M. (2011). Better patients. Better doctors. Better decisions. Envisioning health care 2020. Cambridge: MIT Press.

Hills, T., Stroup, W., & Wilensky, U. (2005). *Patterns of risk seeking and aversion among pre-service teachers: Mathematical decisions, preferences. Efficacy, and participation.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, April 2005.

Hunter, R. (2010). Changing roles and identities in the construction of a mathematical community of inquiry. *Journal of Mathematics Teacher Education, 13*, 397-409.

Hunter, R., & Anthony, G. (2011). Forging mathematical relationships in inquiry-based classrooms with Pasifika students. *Journal of Urban Mathematics Education, 4*(1), 98-119.

Kieran, C., Krainer, K , & Shaughnessy, J. M. (2013). Linking research to practice: Teachers key stakeholders in mathematics education research. In K. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung, (Eds.), *Third International Handbook of Research in Mathematics Education*(pp. 361 392). New York: Springer.

Lampert, M. & Cobb, P. (2003). Communication and language. In J. Kilpatrick, W. G. Martin and D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp 237-249). Reston, VA: National Council of Teachers of Mathematics.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lerman, S. (2006). Socio-cultural research in PME. In A. Gutierrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education* (pp. 347-366). Rotterdam, The Netherlands: Sense Publishers.

Lubienski, S. T. (2007). Research, reform, and equity in U.S. mathematics education. In N. S. Nasir & P. Cobb (eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp.10-23). New York: Teachers College Press.

Martignon, L. & Krauss, S. (2009) Hands on activities with fourth-graders: a tool box of heuristics for decision making and reckoning with risk. *International Electronic Journal for Mathematics Education*, 4 (3), 117-148

Martignon, L., & Kurz-Milcke, E. (2006). Educating children in stochastic modeling: Games with stochastic urns and colored tinker-cubes. *Paper presented at the Seventh International Conference on Teaching Statistics (ICOTS 7),* Salvador, Brazil. International Association of Statistical Education (IASE).

McFeetors, J. & Mason, R. (2005). Voice and success in non-academic mathematics courses: (re)forming identity. *For the Learning of Mathematics, 25* (3), 16-23.

Mercer, N. & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education, 20* (6), 507-528.

Meyer, D. K. & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist,* 37(2), 107-114.

Ministry of Education. (2007). *The New Zealand Curriculum*. Wellington: Learning Media

Nasir, N. S., Hand, V., & Taylor, E. V. (2008). Culture and mathematics: Boundaries between "Cultural" and "Domain" knowledge in the mathematics classroom and beyond. *Review of Research in Education, 32*, 187-240.

National Council of Teachers of Mathematics. (2000). Communication, in *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Governor's Association for Best Practices and Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Retrieved October 1, 2014 from http://www.corestandards.org/

Rohrmann, B. (2005). *Risk Attitude Scales: Concepts, Questionnaires, Utilizations* Project Report University of Melbourne/Australia. Available at www.rohrmannresearch.net

Schlottmann, A., & Wilkening, F. (2011). Judgment and decision making in young Children Probability, expected value, belief updating, heuristics and biases. In M. K. Dhami, A. Schlottmann, & M. Waldmann (Eds.), *Judgment and decision making as a skill: Learning, development, and evolution* (pp. 55-84). Cambridge: Cambridge University Press.

Shapira, M. M., Nattinger, A. B., & McHorney, C. A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*, 21(6), 459-467.

Sharma, S., Doyle, P., Shandil, V., & Talakia'atu, S. (2011). Developing statistical literacy with Year 9 students. *Set: Research Information for Educational Research* 1:43-60.

Silver, E. A., & Smith, M. S. (1996). Building discourse communities in mathematics classrooms. In P. Elliott & M. Kenney (Eds.), *Communication*

Stipek, D,. Salmon, J. M. Givvin, K. B. Kazemi, E. Saxe, G, MacGyvers, V. L. (1998). The Value Convergence) Practices Suggested by Motivation and Promoted by Mathematics Education Reformers. *Journal for Research in Mathematics Education*, 29(4), 465-488.

Streimatter, J. (1997). An exploratory study of risk-taking and attitudes in a girls-only middle school math class. *The Elementary School Journal,* 98(1), 15-26

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science,* 333(6048), 1393-1400.

Stein, M. (2001). Mathematical argumentation: Putting umph into classroom discussions. *Mathematics Teaching in the Middle School, 7* (2), 110.

Streitmatler, J. ((1997). An Exploratory study of risk -taking and attitudes in a girls-only middle school math class. *The Elementary School Journal,* 98(1), 15 – 26.

Sullivan, P. and Mornane, A. (2014). Exploring teachers' use of and students' reactions to challenging mathematics tasks. *Mathematics Education Research Journal,* 26(2), 193 – 214.

Sullivan, P. . Mousley, J., & Zevenbergen, R. (2006). Teacher actions to maximize 122 mathematics learning opportunities in heterogeneous classrooms. *International Journal of Science and Mathematics Education,* 4, 117-143.

Till, C. (2014). Risk literacy: First steps in primary school. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute. iase-web.org [© 2014 ISI/IASE]

Van de Walle, J. A., Bay-Williams, J. M. Lovin, L. H. & Karp. K. S. (2006). *Teaching student-centered mathematics: Developmentally appropriate instruction for grades 6-8 Vol 3, Second edition*, Boston, MA: Pearson Education.

von Glasersfeld, E. (1993). Questions and answers about radical constructivism. In K. Tobin (Ed.), *The practice of constructivism in science education* (pp. 24-38). Washington: American Association for the Advancement of Science.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need – and what we can do about it*. New York, NY: Basic Books.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals.* Mahwah, NJ: Lawrence Erlbaum.

Weber, E. U., Blais, A., & Betz, N. E. (2002) A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors." *Journal of Behavioral Decision Making,* 15, 1-28.

Winsor, M. S. (2007). Bridging the language barrier in mathematics. *Mathematics Teacher*, 101, 372-378.

Whitenack, J. & Yackel, E. (2002). Making mathematical arguments in the primary grades: The importance of explaining and justifying ideas, *Teaching Children Mathematics, 8* (9), 524-528.

Xie, X,. Wang, M., & Xu, L. (2003). What risks are Chinese people concerned about? *Risk Analysis*, 23(4), 685-695.

Pedagogy of Risk:
Why and How Should We Teach Risk in High School Math Classes?

Nenad Radakovic
University of Toronto, Canada

Abstract: Risk is everywhere yet the concept of risk is seldom investigated in high school mathematics. After presenting arguments for teaching risk in the context of high school mathematics, the article describes a case study of teaching risk in two grade 11 classes in Canada- an all-boy independent school (23 boys) and a publicly funded religious school (19 girls and 4 boys). The findings suggest that the students possessed intuitive knowledge that risk of an event should be assessed by both its likelihood and its impact. Following and amending pedagogic model of risk (Levinson, R., Kent, P., Pratt, D., Kapadia, R., & Yogui, C., 2012), the study suggests that pedagogy of risk should include five components: 1) knowledge, beliefs, and values, 2) judgment of impact, 3) judgment of probability, 4) representations, and 5) estimation of risk. These components do not necessarily appear in the instruction or students' decision making in chronological order,; furthermore, they influence each other. The implication for mathematics education is that a meaningful instruction about risk should go beyond mathematical representations and reasoning and include other components of the pedagogy of risk. The article also illustrates the importance of reasoning about rational numbers (rates, ratios, and fractions) and their critical interpretation in the pedagogy of risk.

Keywords: probability teaching and learning, risk literacy, teaching and learning of risk, inquiry-based learning, understanding of rational numbers; risk estimation.

Introduction

Risk as a concept permeates all aspects of our society, it appears in every activity we do as humans and it shows in diverse disciplines such as mathematics, physics, engineering, sociology, and psychology. Yet, despite its importance, it is rarely included in school mathematics. In this article, I explore teaching risk in the context of high school mathematics.

There is a consensus amongst experts that most people are unable to adequately interpret and communicate risk (Kahneman, Slovic, and Tversky, 1982; Rothman, Montori, Cherrington, A., & Pigone, 2008). The problem of improving understanding of risk has been addressed in the specific context of public health and financial counselling, yet it has only begun to be explored within educational research (Pratt et al., 2011). Despite a recognized and urgent need for risk education, there is a lack of agreement on its definition. The concept of risk exists at the intersection of many related fields—mathematical, health, statistical, probability, scientific, and financial, among others. In this study, I will situate risk within the fields of statistical and probability literacy as these fields focus on uncertainty and chance, both important elements of risk-based reasoning. Most current approaches to literacy recognize it as more than a minimal subset of content knowledge in a particular field (see, for example, Gal, 2004a). Further, the definition of literacy has been expanded to include "desired beliefs, habits of mind, or attitudes, as well as a general awareness and a critical perspective" (Gal, 2004a, p. 48). Consistent with Gal's (2004a, 2004b, 2005) research on statistical and probability literacy, I define pedagogy of risk to include knowledge (e.g., probability content knowledge) and dispositional elements

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 307-329

(e.g., beliefs and attitudes about risk) as I examine its place within the secondary mathematics curriculum.

Researchers and policy makers have recognized the need for education about risk (Gigerenzer, 2002; Kolsto, 2001; Levinson, R., Kent, P., Pratt, D., Kapadia, R., & Yogui, C., 2012; Pratt et al., 2011). Pratt et al. (2011) provide examples from the UK curricular documents which call for teachers to teach probability through situations involving risk. In the context of the Ontario secondary school curriculum, the concept of risk can be found in multiple subject areas, including science (e.g., students are expected to learn to analyze risk of introducing particular technology to ecosystems), physical and health education (e.g., risk involved in participation in a physical activity), and family studies (e.g., risk of contamination in food) (Ontario Ministry of Education, 2011). Given that understanding of risk includes a strong quantitative component, the mathematics classroom is an appropriate setting for the exploration of its pedagogy. However, in Ontario, mathematics curriculum documents do not focus on risk. Moreover, throughout the secondary mathematics curriculum, there are limited mentions of risk; these fall within the context of financial mathematics as well as the promotion of students' risk taking which is considered "necessary to become successful problem solvers" (Ontario Ministry of Education, 2005, p. 24). The exploration of the pedagogy of risk has begun only recently. The most comprehensive research in pedagogy of risk was done by the researcher involved in the Institute of Education's TURS Project (Promoting Teachers' Understanding of Risk in Socio-scientific Issues). The research, which involved in-service teachers, problematized risk education as the interplay of mathematical knowledge, beliefs, context, and content knowledge.

Despite calls for teaching risk in the classroom, and despite the explorations by the TURS research group, there remains a lack of research in the mathematics classroom setting and involving students. The purpose of this study is to address the lack of research by exploring the ways in which risk could be taught within the mathematics classroom. Specifically, this study explores the ways that secondary school mathematics instruction can support students' developing understanding of risk, and I focus on the following guiding question, namely how do secondary school students reason and make decisions about risk?

Review of Literature

The Concept of Risk

Risk is a concept that is prevalent in many disciplines and the term 'risk' has been used in many distinct yet connected ways. Hansson (2009) distinguishes between five different definitions of risk: 1) risk as an unwanted event which may or may not occur; 2) the cause of an unwanted event which may or may not occur; 3) the probability of an unwanted event which may or may not occur; 4) the fact that a decision is made under conditions of known probabilities; and 5) the statistical expectation value of unwanted events which may or may not occur.

The third, fourth, and fifth definitions are the most common in mathematics. The third definition aligns with the view that a risk associated with an event is a quantifiable uncertainty (Gigerenzer, 2002), which is equivalent to the likelihood or probability of the event. This definition of risk is suitable when the events have similar consequences, but it becomes problematic if the impact of each event is different. For example, the likelihood of a person catching a cold is relatively large but its impact on the person's life is most likely to be minimal, whereas the likelihood of getting killed in a terrorist attack is relatively small but the impact is immense. In order to account for both likelihood and impact, proper understanding of risk requires the coordination between judgments of probability and impact (Pratt et al., 2011), which corresponds to the fifth definition, the statistical expectation. This coordination can be done informally, but also formally using mathematical representations. Symbolically, the fifth definition of risk can be written as

$$R = \sum_{i=1}^{n} p_i d_i$$

where the overall risk, R, of a hazard, is the sum of the products of the probability (p) and disutility or impact (d) of each event associated with the hazard (Pratt et al., 2011). For example, to assess the overall financial risk of owning a car, one would find the probability of each outcome (e.g., flat tire), multiply those by the financial impact, and then obtain the total sum of all the products. The approach based on the above formula is known as the utility theory of risk (Levinson et al., 2012) and is the standard approach in technical risk analysis (Moller, 2013).

Cultural Perspectives

Utility theory and technical risk analysis are not the only approaches to risk. Technical risk analysis, which is a domain of philosophy, statistics, and economics, has been extended to risk governance which involves actors' understanding and handling of risk (Lidskog & Sundqvist, 2013). However, risk governance is a complex task, particularly in the case of global risks such as terrorism, catastrophic weather due to climate change, financial meltdown, and nuclear accidents such as the radiation leakage due to the Fukushima nuclear disaster. The anticipation of global risks can seldom be determined using methods of science. The less we are able to calculate risk, the more the balance shifts toward the cultural perspectives on risk (Beck, 2009). It follows that assessing risk goes beyond the utility theory. Assessing risk in the vast majority of social situations involves more than individual considerations of maximizing utility; rather, it is a dynamic consensus-making political process involving diverse actors and contexts (Douglas, 1992). Consistent with the cultural perspective on risk, "sociology opposes any kind of reification of risks, in which risks are lifted out of their social context and dealt with as something uninfluenced by the activities, technologies, and instruments that serve to map them" (Lidskog & Sundqvist, 2013, p. 77).

Beyond Utility Theory and Cultural Theories. Reality is "neither reducible to something out there, beyond human action, nor reducible to something in there, to human thoughts and actions" (Lidskog & Sundqkvist, 2013, p. 98). Reification of risk—the belief that risk is void of social context—is problematic. However, the social purification of risk the notion that risk is just the product of social factors (Lidskog & Sundqvist, 2013)—is also problematic. The question, then, is how to reconcile naïve realism and idealism. The third way should not be obtained by the combination of constructivism and realism. Instead, the focus should be on "the dynamic interplay between different factors that make up reality" (Irwin & Michael, 2003; Latour, 1993, 2004, 2005). Latour suggests that one can transcend the perceived dichotomy between utility and cultural theories by focusing on the production of risk:

> Risks are produced by practices, by actors using instruments and technologies. It is therefore misleading as a sociologist to focus on perceptions, opinions and experience. Instead, the focal point for sociology should be to explore how risks are produced, by what means and with what effects. (Lidskog & Sundqvist, 2013, p. 99)

Recently, technical risk analysis has recognized the importance of cultural theory and has treated values (individual or collective) on par with empirical data. For example, the structured decision-making approach (Gregory et al., 2012) takes into account the complexity of environmental decision making by considering uncertain science and multiple stakeholders' values and priorities.

In addition, the precautionary principle (Wiener, 2002) can serve as a mechanism to support decision making when there is a lack of scientific evidence. The precautionary principle states that "[w]hen an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically" (Science and Environmental Health Network, 1998).

Teaching and Learning of Risk. The tension between the utility and cultural theories is reproduced in the teaching and learning of risk. In the following section, I discuss the two predominant approaches to the teaching and learning of risk: deficit theory of risk (utility approach) and pedagogical approach (Levinson et al., 2012).

Deficit Theory of Risk

If we accept the assumption of utility theory that there is such thing as objective or actual risk, then the goal of risk education is to evaluate the learner's perceived risk and, through cognitive adjustment, to align it with actual risk (Levinson et al., 2012; Lidskog & Sundqvist, 2013). The deficit theory has been successful in solving problems that are well defined—in other words, the problems for which there is a unique solution, such as a subset of problems involving Bayesian inference (Gigerenzer, 2002; Kahneman et al., 1982).

Research on students' reasoning suggests the importance of different representations in solving Bayesian problems. For example, Zhu and Gigerenzer (2006) propose that students' abilities to solve Bayesian tasks vary depending on the data being represented in terms of natural frequencies or probability. When the binomial hypothesis problem was presented in terms of probability to fourth, fifth, and sixth grade students, none of the students were able to estimate the Bayesian posterior probability. On the other hand, when the same information was presented as a natural frequency (e.g., rather than represent the probability as 0.07, state that "out of 100 women, 7 have cancer"), reasoning about conditional probability showed a steady increase from fourth to sixth grade, reaching an average level of 19%, 39%, and 53%, respectively, in two studies.

In summary, researchers who apply the deficit theory of risk education begin with the analysis of students' perceptions (heuristics and biases) and then propose an intervention (e.g., the use of natural frequencies) in order to align the perceived risk with the actual risk. (Figure 1)

Critiques of the Deficit Theory of Risk. Simple Bayesian tasks can be approached using the deficit model because, within the context of the problem, there is an actual risk—or, to be more precise, an actual quantifiable risk. However, if the question is reframed as "How often should women over forty get mammograms?" or, even more specifically, "Should a particular person get a mammogram?", it becomes ill-defined, meaning there is neither a clear solution nor a clear method of arriving at a solution. The deficit model is particularly inadequate when dealing with technoscientific situations such as epidemics, climate change, energy policy, or pharmaceutical research, where there is no consensus on how probability or impact should be calculated (Levinson et al., 2012; Lidskog & Sundqvist, 2013).

Even when problems are well defined, it is debatable whether individuals base their decisions simply on cognitive heuristics. In other words, when making risk-based decisions, are we really performing calculations at a varying level of complexity depending on our background in risk analysis? Paul Slovic, a veteran of heuristics research, began to doubt this several decades ago:

> I recall, in the midst of this growing collection of heuristic strategies, wondering how people decided when it was safe to cross a busy street. Certainly they were not calculating probabilities and utilities or their summed products, and the known judgment heuristics did not seem to offer any insight. (Slovic, 2010, p. xx)

He instead proposes that any risk decision includes the affect heuristic, which is a cognitive process in which individuals use feeling (positive or negative) as a guide to evaluating risk.

Deficit theory also does not take into account that risk situations are "constructed by different histories, narratives, and experiences" (Levinson et al., 2012, p. 216). For example, Levinson and Rodd (2009) investigated pre-service teachers' conceptions of risk related to the question of whether malaria is a major risk in travelling to West Africa. A student who had had an experience with malaria downplayed the risk; "what was seen as a major risk by one person was not perceived as a significant risk by another" (Levinson et al., 2012, p. 216).

Another problem with deficit theory is that it places expert knowledge before the knowledge of laypeople, regarding them as "poorly informed in comparison to the 'precise' and 'scientific' analyses of experts" (Beck, 2009, p. 12). Laypeople, however, "have the competence to contribute to discussions and decisions on risks since they concern them much more than scientific facts" (Lidskog & Sundqvist, 2013, p.94). Levinson et al. (2012) assert that "evolving models of interactions between experts and publics point toward a more reflexive expert perception of public concerns and a realization of the importance of public engagement" (pp. 216-217). Beyond public engagement, Gregory et al. (2012) call for the meaningful inclusion of public knowledge into decision-making processes, focusing in particular on local and traditional knowledge characterized by the reliance on experience and observations rather than experimentation, often expressed in more holistic rather than reductionist fashion, and dealing with particular concerns and context-dependent situations.

Consistent with the value of public knowledge and decision making, Levinson et al. (2012) consider personal models in understanding of risk for two reasons:

> learning involves the modification of preexisting personal models in interaction with others, rather than learning being a process of replacing learners "wrong" thinking with models of "right" thinking and (2) it is critical to respect personal models because personal values (as expressions of personal preferences and ethical positions) and social and affective values are inextricable from making decisions. (p. 217)

It follows that risk should be taught and learned in an environment that creates opportunities "to make explicit values, experiences, and representations of those experiences and probabilities that foreground the decision-making process, and where probabilities can be judged in light of, and interact with, expressed values" (Levinson et al., 2012, p. 228). An inquiry-based approach can offer such an environment (Pratt & Yogui, 2010).

Pedagogic Model of Risk

For the purpose of exploring the pedagogy of risk, researchers involved in the Institute of Education's TURS Project (Promoting Teachers' Understanding of Risk in Socio-scientific Issues) developed a computer microworld called Deborah's Dilemma (Levinson et al., 2011; Levinson et al., 2012; Pratt et al., 2011). In Deborah's Dilemma, students were engaged in a narrative involving a fictitious person, Deborah, who suffers from a spinal cord condition. Based on the data about the side effects of a surgery and the consequences of not having the surgery, pairs of math and science teachers had to choose the best possible course of action for Deborah. One of the outcomes of the research program was the development of the pedagogic model of risk (Levinson et al., 2012).

According to this model, probabilistic judgments lead to the estimation of risk but the judgments are informed by values, experiences, personal and social commitments, as well as representations (see Figure 2). This is in contrast with the utility model of risk, where values are separate from the probabilistic judgments and may only play a role in risk management (following an analysis of risk).

Relevant findings from the study have been used throughout this literature review to outline the elements of the pedagogy of risk.

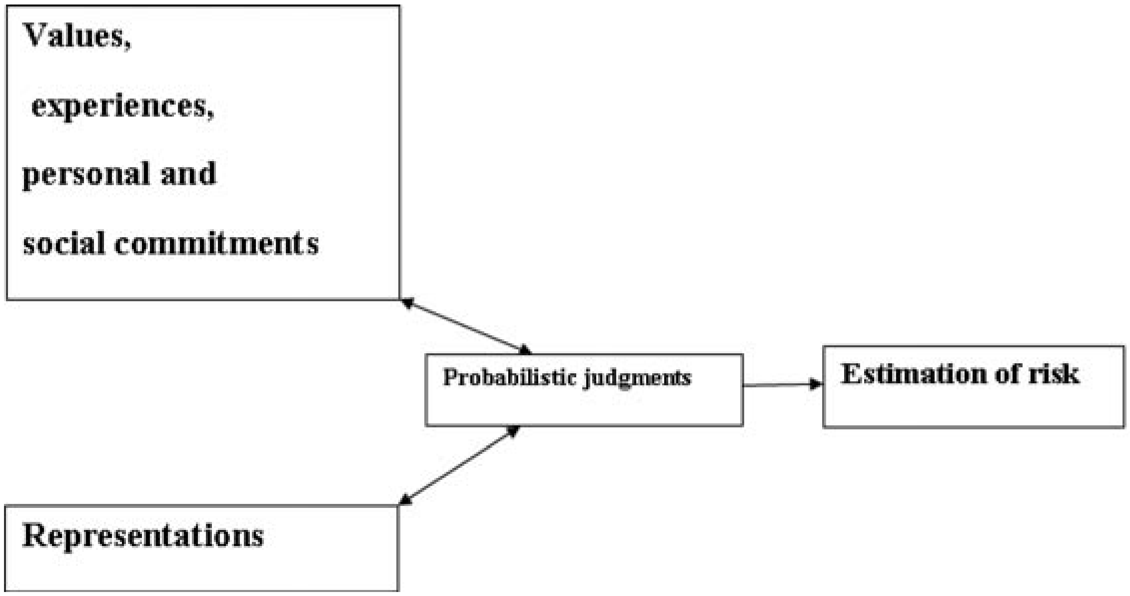


Figure 1. Pedagogic model of risk. Adapted from Levinson, Kent, Pratt, Kapadia, and Yogui (2012).

Elements of the Pedagogic Model of Risk

In this section, I present the existing research that supports and expands upon the Levinson et al. (2012) pedagogic model of risk. This will be followed by research on the estimation of risk, with a particular focus on coordination between probability and impact. Finally, the role of context in the pedagogy of risk will be discussed.

Probabilistic Judgments. In terms of probability knowledge, there are several frameworks that describe probabilistic reasoning. The major frameworks mentioned in the literature are Core Domain of Probability Concepts (Moore, 1990), Probability Thinking Framework (Jones, Langrall, Thornton, & Mogill, 1997; Polaki, Lefoka, & Jones, 2000), and Gal's knowledge elements of probability literacy (Gal, 2004b, 2005). Moore (1990) describes the conditions that students need to satisfy in order to be able to move towards more difficult concepts such as conditional probability. These conditions are: 1) learning to discern the overall pattern of events and not attempt a causal explanation of each outcome; 2) recognizing the stability of long-run frequencies; 3) assigning probabilities to finite sets of outcomes and comparing observed proportions to these probabilities; 4) overcoming the tendency to believe that the regularity described by probability applies to short sequences of random outcomes; and 5) applying an understanding of proportions to construct a math model of probability and develop an understanding of some "basic laws or axioms that include the addition rules for disjoint sets" (Moore, 1990, p. 120).

One of the most comprehensive frameworks describing probabilistic reasoning is provided by Jones et al. (1997). The framework is divided into four constructs: 1) sample space, 2) probability of an event, 3) probability comparisons, and 4) conditional probability. The framework also recognizes four levels of reasoning: 1) subjective, 2) transitional between subjective and naïve quantitative, 3) informal quantitative, and 4) numerical.

Polaki et al. (2000) provide an extension of the Jones et al. (1997) framework. Their framework describes probabilistic thinking across five constructs: 1) sample space, 2) probability of an event, 3) probability comparisons, 4) conditional probability, and 5) independence. The Polaki et al. (2000) framework has been used to explain requirements for understanding compound events in terms of the sample space and the probability of the event constructs (Nilsson, 2007). The Jones et al. (1997) framework, which serves as a basis for the Polaki et al. (2000) framework, was used by Langrall and Mooney (2005) to interpret the grade three students' understanding of probability in Falk and Wilkening's (1998) research.

Representations. One of the biggest challenges for learners of risk is the issue of representations. The claim that the Bayesian problem is simpler when the probability is represented in terms of natural frequencies is made elsewhere (Cosmides & Tooby, 1996; Gigerenzer, 1994; Gigerenzer & Hoffrage, 1995). These researchers offer an evolutionary psychology explanation of why most people find Bayesian tasks represented using natural frequencies easier than others. According to this explanation, humans are "hard wired" to deal with natural frequencies rather than with probabilities.

In the Levinson et al.'s (2012) study, however, the choice of representation seemed to depend on the context. For example, when deciding whether Deborah should have surgery based on data about the success rate, a teacher claimed that, if they were arguing for the surgery, they would represent the rate of failure in terms of probability ("less than a half of seventh"), whereas, if the recommendation was against the surgery, they would represent the data in terms of natural frequency ("four people") because it contains information on actual people impacted by the failure. The idea that natural frequencies carry more weight in representing impact is confirmed by Tim, another teacher in the Levinson et al.'s (2012) study:

> I mean if you're going to say "60 people died from this procedure," is that enough to tempt someone to say "alright I'll give that a go"? Ok that would look bad because 60 people is more impacting on you than one in 1000, one in 10,000. Those big figures will convince you, but I think "60 people died from this last year" convinces you in a different way, even though the figures, you know that's where the one in 50,000 comes from. The way you present your data is very important to an individual. (p. 223)

Levinson et al. (2012) also documented that teachers made mistakes in calculating percentages:

> In paired dialog, it was easy for teachers to miscalculate small percentage values into figures and proportions more commonly used in everyday discourse. Overall, this illustrates a common problem, where people find large numbers and low probabilities difficult to comprehend. It suggests the need to take care in designing materials about risk, possibly highlighting the need to support students and consumers in negotiating and interpreting the ways in which probabilities are represented. (p. 223)

In terms of representations of impact, Pratt et al. (2011) document "fuzzy qualitative descriptors" used by students to roughly calculate impact ("serious," "massive," "bad," "fine," "big") (p. 339). Qualitative representations of impact (outcomes) were similarly documented by the Levinson et al. (2012) study; all three pairs of teachers considered impact in deciding whether Deborah should have the operation using phrases such as "impact on her life," "pain threshold," and "prohibitively dangerous option" to describe impact.

In contrast with qualitative representations of impact, there is a lack of research on what quantification of impact looks like in the classroom. There is, however, empirical evidence that quantification does not come easy. Pratt et al. (2011) state that there was no opportunity for teachers to quantify impact:

> At no point did any of the six teachers attempt to numerically quantify impact and certainty [sic] in designing the task we provided no explicit encouragement to move beyond fuzzy quantifications of how much additional pain might be caused by a lifestyle decision. Perhaps because of the difficulties in quantifying impact, and to some extent of making sense of the odds, we observed no attempts to quantify risk per se in a formal way. (p. 339)

We can conjecture, however, that quantitative reasoning about impact will be consistent with quantitative reasoning described by statistical literacy frameworks. According to these frameworks, understanding of rational numbers, including the understanding of ratio and proportion, is another domain of quantitative knowledge and is a necessary prerequisite for performing at the highest level of statistical understanding (Watson & Callingham, 2003; Watson & Shaughnessy, 2004). According to Watson and Callingham (2003), statistical literacy is a hierarchical construct with proportional reasoning featured at its highest levels (see Table 1). At the highest level, critical mathematical, students are required to use "proportional reasoning associated with ratio and appropriate part-whole interpretations, the ability to use rates in calculating costs" (p. 18).

Proportional reasoning causes difficulty for many middle school (Lamon, 2007) and high school students. For example, Akatugba and Wallace (1999) studied students' proportional reasoning in a physics class and found that students had difficulty performing mathematical operations that were not explicitly stated in the task. The conclusion of the study was that students' understanding of mathematical processes involved in proportional reasoning was inadequate.

Values, Experiences, Personal, and Social Commitments. Students also must be aware that the collection, generation, and interpretation of data are influenced by social factors and are consequently value-laden (Pratt et al., 2011; Watson, 1997). In order to examine the social factors, students must be able to critically examine data (Gal, 2004a; Pratt et al., 2011). In addition, thinking about risk also involves decision making that can function on an individual level, societal level, and within the intersection of these levels (Pratt et al, 2011).

The importance of critical evaluation of data is consistent with the consensus among educators stating that literacy should extend beyond a minimal subset of skills expected for all. According to Gal (2005):

> increasingly the term literacy, when used as part of the description of people's capacity for goal-oriented behavior in a specific domain, suggests a broad cluster not only of factual knowledge and certain formal and informal skills, but also a desired beliefs, habits of mind, or attitudes, as well as a general awareness and a critical perspective. (p. 42)

For the purpose of the critical evaluation of data, Gal (2004a) introduced a list of critical or worry questions which enable individuals to critically evaluate information provided. The critical questions include questions about reliability and the validity of data (Gal, 2004a, 2004b, 2005).

The students' perception of reliability of data is an important critical element. There is evidence that some individuals associate reliability with disinterestedness. For example, Pratt et al. (2011) found that two participants in their study (Linda and Adrian) were suspicious of the surgeon who they thought may be "drumming up the business" (p. 340).

Levinson et al. (2012) found that students "recognize the role of trust and authority in giving meaning to the data" (p. 224). The authors reported that a participant (Linda) found the spine doctor's recommendation reliable because he "knows more about it than the other people, and he's seen more of these people" (Levinson et al., 2012, p. 224). The notion of trust in groups that are involved in measurement is evident in Kolsto's (2001) study. Despite the fact that the students were sceptical of the power company's risk evaluation and the power company was marked as interested, the students accepted the company's claims concerning magnetic field strength from different sources.

Besides the issues of reliability, there is an issue of content validity, which includes the questions of how the statement about data was derived, whether the claims are supported by the data, and whether additional information and interpretations are needed (Gal, 2004b). Kolsto (2001) provides evidence that students are quite vigorous in questioning the source of data but are less likely to question the validity which requires students to analyze the content of the source and evaluate the arguments presented in the document.

Research on affect heuristics also sheds light on the ways that individuals make decisions about risk. Slovic, Finucane, Peters, and MacGregor (2010), influenced by the dual processes theory (Kahneman & Frederick, 2002), suggest that human reasoning about risk consists of two cognitive systems: one is the experiential, intuitive system that helps us make quick assessments about the safety of a situation ("a gut feeling"), the other is the analytic system that helps us evaluate our thinking. Slovic et al. (2010) do not want to fall into a trap of deficit theorists by favouring the analytic system over the experiential. Instead, they use empirical evidence to argue that affect that stems from the experiential system helps us to make decisions quickly in an uncertain and dangerous world:

> We now recognize that the experiential mode of thinking and the analytic mode of thinking are continually active, interacting in what we have characterized as the 'dance of affect and reason' (Finucane et al., 2003). While we may be able to 'do the right thing' without analysis (e.g. dodge a falling object), it is unlikely that we can employ analytic thinking rationally without guidance from affect somewhere along the line. Affect is essential to rational action. (Slovic, Finucane, Peters, & MacGregor, 2010, p. 24)

An example of affect heuristics is the feeling of dread (Fischhoff et al., 1978), which has been shown to be a major predictor of public perception of risk for a wide range of phenomena (Slovic et al., 2010). For example, the feeling of dread towards nuclear power has resulted in the view of nuclear power "as a technology whose risks are uncontrollable, lethal, and potentially catastrophic" (Slovic, Fischhoff, & Lichtenstein, 1982, p. 485).

Estimation of Risk. In contrast to the research on probability, there are no comprehensive frameworks for analyzing students' estimation of risk, particularly the estimation that involves coordination between probability and impact. However, the research of Pratt et al. (2011) suggests that, in order to comprehend the coordination, students must have both algebraic and analytic geometry skills that will enable them to manipulate formulae and graph functional relationship between two variables.

Levinson et al. (2012) state that the participants in their study did not sufficiently coordinate between probability and impact; instead, the majority of discussion was based on personal attitudes of the group of teachers. The authors state that "while there was some discussion of probabilities, these only interacted with the decision on outcomes in a marginal way or provided insufficient background for a decision to be made" (p. 226). In addition, Levinson et al. state that:

> there were relatively few instances where the teachers simultaneously balanced changes in lifestyle against the likelihood of the operation resulting in serious harm. This might have been a problem in the way the data were presented, but it was more likely that personal preferences were driving decisions irrespective of emerging evidence. (p. 228)

The Role of Context and Content Knowledge in Pedagogy of Risk

Pratt et al. (2011) ask whether context may impede students' understanding of risk, drawing on examples from the previous studies which suggest that context may be detrimental to the mathematical understanding of risk. Pratt et al. (2011) conclude that the understanding of risk is closer to statistics than mathematics and that the context is crucial. If we strip away context and reduce the task of assessing risk to the mathematical coordination between likelihood and impact, we can see that the meaning is lost. Also, the numbers (quantitative data) have to be viewed in context. Pratt et al. (2011) also consider the issue of who the decision maker is as an element of context. In other words, students will respond differently depending on whether they are making decisions about themselves or another person.

Data concerning risk and uncertainty are never decontextualized and comprehensive interpretation of data involving risk requires placing data in an appropriate context (Gal, 2005). Therefore, contextual knowledge is essential for risk analysis. For example, the analysis of risk of a nuclear power plant accident requires content knowledge about nuclear power plants. Understanding of risk contains many elements of mathematical reasoning (logical, probabilistic, and proportional reasoning); it also contains many elements of the quantitative reasoning in context (Mayes, Peterson, & Bonilla, 2012). According to the authors:

> Quantitative reasoning in context (QRC) is mathematics and statistics applied in real-life, authentic situations that impact an individual's life as a constructive, concerned, and reflective citizen. QRC problems are context dependent, interdisciplinary, open-ended tasks that require critical thinking and the capacity to communicate a course of action. (p. 10)

Pratt et al. (2011) acknowledge, based on their study of math and science teachers, that reasoning about risk is highly contextualized. They differentiate between the context of the problem and the setting where reasoning takes place, namely reading, inquiry, and pedagogic setting (Monteiro & Ainley, 2003). An example of a reading setting is reading newspaper articles or advertisements, an inquiry setting is one in which an individual engages with data to solve the problem, and a pedagogic setting is the school setting where highly formal and mathematical ways to solve very specific problems are used (Pratt et al., 2011). It can be argued that, when solving problems, students can have different goals within each context.

Pratt et al. (2011) conjecture and show evidence that these three settings differ in the way participants use cognitive resources. They assert that cognitive resources most readily used in the reading setting are affective (emotional) responses followed by the understanding of context. In the inquiry setting, understanding of the problem context takes precedence over mathematical and statistical knowledge, though this knowledge is also important. Finally, the authors conjecture that, in the pedagogic setting, which is intended to teach and assess particular ideas, statistical and mathematical ideas are prioritized, whereas affective resources are less likely to be drawn upon.

One of the prerequisites for understanding of risk is content knowledge (e.g., in order to be familiar with risk of nuclear power plants accidents, we should have relevant knowledge about nuclear power). However, possession of content knowledge does not guarantee that the knowledge will be used in risk assessment. For example, Levinson et al. (2012) state that:

> there were many opportunities in the microworld for the teachers…to make use of relevant scientific knowledge in helping to evaluate risk, but none chose to do so,

> reflecting other accounts, where scientific knowledge and information are either marginal or irrelevant to lay decision making. (p. 228)

Teaching Risk in the Classroom

There is a lack of classroom studies of students' understanding of risk; the studies that do exist stress the importance of treating risk-based decision making as a complex enterprise. The importance of the complexity of decision making is highlighted by Monteiro and Ainley (2003) in their study of student teachers who drew on four distinct resources: mathematical knowledge, contextual knowledge, affective responses, and personal experiences. The authors found that "if attention is focused exclusively on one of these sources, then the judgment may be distorted" (as cited in Pratt et al., 2011, p. 338).

In order to create a classroom environment conducive to the complex view of risk, Kolsto (2006) suggests that students need "easy access to an appropriate range of information and viewpoints" (p. 1711). For example, in Kolsto's study of high school student decision making related to electrical power lines, many students were only drawing their conclusions based on research-related information. Kolsto (2006), citing Aikenhead (1985), suggests that if we want students to draw from wider domains (including values), we need to include tasks in which students are confronted with this information.

Summary: Pedagogy of Risk

In summary, a comprehensive pedagogy of risk should embrace complexity of the concept of risk as well as human risk-based decision making. Research supports the claim that classroom instruction needs to be inquiry-based and embrace students' experiences, beliefs, and values (Pratt et al., 2011; Levinson et al., 2012). Based on students' risk-based reasoning, the goal of the pedagogy of risk ought to be the articulation of pedagogical strategies needed for teaching and learning and finding a place for the instruction within the mathematics curriculum. However, in order to articulate the pedagogical strategies, there ought to be more research on how students reason and make decisions about risk in the mathematics classroom setting.

Methods

I applied a qualitative case study approach as I explored the teaching of risk in two 11th grade classrooms that were using an inquiry-based learning approach to pedagogy. This section begins with a justification of my selection of a qualitative case study methodology for conducting research in the classroom, followed by my reasoning for the use of inquiry-based learning. I then outline the selection of the school, teachers, and participants, as well as the chronology of research, including the initial interviews with teachers, initial assessment of students, inquiry-based activities, final assessment, and final interviews with teachers. The section concludes with a detailed description of methods used for data collection and analysis, and also the ethical considerations relevant to my research.

Research Setting and Participants

The first research setting was Dale Academy, an all-boys private secondary school following the International Baccalaureate curriculum. Every student at the Dale Academy had access to many educational resources, including laptop computers and wireless Internet. The reason why this school was chosen was to be able to see how risk pedagogy can be approached in settings in which there is no lack of resources. The 11th grade class was chosen because the International Baccalaureate probability and statistics unit was a good place for teaching and learning about risk. In order to have a greater diversity of participants, the second school was St. Hubertus Secondary School, a co-educational school with no direct access to laptops and no wireless Internet access. I did most of the teaching in the study—the two teachers, Breanna and Clarissa, were there to help me plan the lessons, observe them, and assist me with the logistics and classroom management. Thus, the case study centres on the students and me, whereas the role of the classroom teacher was not explored in the study.

Inquiry-Based Learning Approach

In this section, I describe the inquiry-based learning approach (IL) and explain why it is appropriate for my study. As mentioned in the literature review, students' risk-based reasoning should be rich and complex and involve not only thinking about mathematics and content knowledge but also personal beliefs and experiences (Pratt et al., 2011). IL is consistent with the above statement because, in IL, "students learn content as well as discipline-specific reasoning skills and practices (often in scientific disciplines) by collaboratively engaging in investigations" (Hmelo-Silver, Duncan, & Chinn, 2007, p. 100). Inquiry-based learning involves authentic tasks that enable students to engage with the topic of inquiry. There are different views on what makes a task authentic (Levinson et al., 2012), but for the purpose of this thesis, I follow the distinction made by Murphy et al. (2006) between cultural authenticity and personal authenticity. They assert that cultural authenticity is present when students engage in a common social issue, whereas personal authenticity refers to an issue of importance to an individual student.

I draw upon the 5 E Model (Bybee et al., 2006) in which IL consists of five steps: 1) engage, 2) explore, 3) explain, 4) elaborate, and 5) evaluate. While presented linearly, the five steps do not always proceed chronologically and each of them may contain the teacher's help and guidance. The engagement stage consists of an activity or activities in which students engage in an issue which is either socially or personally authentic. In the exploration stage, students collaborate through an authentic task and begin to clarify their understanding of major concepts. Students are involved in the explanation stage when they construct concepts and processes about which they are exploring and learning. Finally, the elaboration activities challenge students to apply what they have learned to a new situation, and evaluation involves both students' and teachers' assessment of progress for the purposes of informing instruction (Bybee, et al., 2006).

The IL module started with an initial assessment of the student understanding of risk and knowledge of the Fukushima accident in order to engage students in the issue. The initial individual written assessment consisted of two questions setting the context of the issue. The first question was:

> Do you agree or disagree with the following statement? Explain your reasoning.
>
> "There have been around 500 nuclear power plants built since the 1950s. So far there have been only three significant nuclear power plant accidents. This makes nuclear power relatively safe compared to other means of generating power."

The purpose of the initial assessment was to show evidence that students possessed knowledge about impact prior to the instruction and to investigate language the students used when talking about impact. In addition, I wanted to investigate whether there was any evidence for the coordination between likelihood and impact as discussed by Pratt et al. (2011).

The second question consisted of two parts. First, the students were asked to write a paragraph on what they know about the accident. Their responses were then marked on a scale of 0 to 3 corresponding to the content of their answer (as described in the data analysis section as well as the findings). The second part of the question was: "What is your opinion on safety of nuclear power plants and how did the Fukushima incident influence your thinking?" The purpose of the question was to give students more opportunity to consider reasoning about risk in terms of probability and impact as well as to gauge their personal values, experiences, and personal and social commitments. The initial assessment was done during class and took around 30 minutes to complete. Following the initial assessment, there were 3 hours and 45 minutes of instruction (over the three 75-minute periods) on determining the empirical probability of nuclear power plant accidents. The activity started with a 75-minute lecture that defined key terms: probability, theoretical probability, and empirical probability.

Following the introduction, there was a 75-minute group activity (full period) with the following objectives: 1) to critically evaluate the sources of data provided; and 2) to estimate the empirical probability. The activity was specifically designed to contain the "explore," "explain," and "elaborate" elements of the 5 Es. The students were given a worksheet in which they were presented with four websites to use as potential data sources. The websites contained nuclear power plant accident data and could be easily accessed with laptops. The websites were:

•World Nuclear Association (WNA, 2011) – an organization representing the interests of the nuclear energy industry

•Greenpeace (Greenpeace, 2010)– an environmental activist group, giving a comprehensive list of incidents and accidents

•Datablog (Rogers, 2011) – the statistics blog from the *Guardian*

•Ecocentric Blog (Harrel, 2011) – the environmental blog from *Time* magazine

I chose to give students specific websites rather than having them freely explore the Internet because by selecting from a shared set of websites, I was able to gain insight into the reasons why they picked one over the other.

After deciding which website to draw data from, students were instructed to estimate the probability of a nuclear power plant accident. Following the activity, the groups presented their findings to the class; this took approximately 75 minutes (one full period). The group presentations had the potential to contain the "explore," "explain," "elaborate," and "evaluate" elements of the 5 Es.

Following the first activity, the second group activity (also 75 minutes) was completed and involved interpretation of data including likelihood and impact. The objective of the activity was to introduce students to the assessment of the impact, both qualitative and quantitative, and the coordination of the likelihood and impact, and to present them with the idea that the assessment may be value-laden. In terms of the five Es proposed by Bybee et al. (2006), it contained the elements of "explore," "explain," and "elaborate." This was followed by the presentation of data (75 minutes), containing the elements of the last four Es. Throughout the group activities, I was a facilitator assisting in student learning and using direct instruction to clarify certain points—the direct instruction was given to either the groups or the whole class.

Data Collection

Data collection for the first school started in March of 2011, with the four half-hour semi-structured interviews with the teacher in order to prepare for the lessons. These interviews were audio-recorded and helped me plan the activities in terms of choosing an appropriate class, duration of the activities, and the time period in which I could do the data collection. This data informed the IL activities but was not directly used in data analysis. All of the lessons were video-recorded; a camera captured the whole class and each group was video-recorded. In addition, student-written work from the initial assessment questions, handouts, and the construction paper given to each group for the activities was collected. I also kept field notes during the activities. This was challenging since I was also facilitating the lessons, and did not result in a rich and consistent data source. However, after each lesson, I recorded a reflection on each lesson and a short (5 minutes) debriefing with the teacher that was audio-recorded. The debriefing consisted of the teachers' assessment of the lesson, mostly their comments on the student engagement and on the logistics for the next activity. This data was used to make conjectures about the difference in engagement between the two groups. At the culmination of the series of lessons, I had an hour-long semi-structured interview with each teacher about her reflection on the activities as well as on her background as a teacher. These data were used to create a background of the teacher and also to make conjectures about the engagement differences between two groups. Preparation for the data collection at Dale Academy began in March of 2011. The data collection began and ended in May of the same year. Preparation for the data collection at St. Hubertus Secondary School began in early September 2011. The second data collection, at St. Hubertus, began in October and ended in November of 2011 and had the same structure as the first data collection.

Findings and Discussion

The findings suggest that the source of data (the four websites) was chosen based on reliability and the visual presentation of data. The judgment about reliability was based on reputation, sense of disinterestedness, neutrality, caution towards Internet documents, and trust in authority. In both settings, students found the *Guardian* website most trustworthy, being a "reputable newspaper." For many students, a source was reliable if it was seen to be disinterested and neutral. Specifically, for the Dale Academy students, the *Guardian* was considered the website that was not a "stakeholder in the issue" (David, Fahad, Samir, and Sasha) whereas the Greenpeace site was deemed unreliable because it had a clear agenda which was seen as anti-nuclear. Interestingly, the World Nuclear Association (WNA), a pro-nuclear power advocate, was not seen as a stakeholder, possibly because the students were not aware of its agenda since they were never explicitly given any information about it. On the other hand, the students at St. Hubertus were given brief descriptions of the sources and the WNA entry specifically stated that WNA was sponsored by the nuclear industry. The students unanimously found WNA to be unreliable. For example, Chloe, Mina, Larissa, Dana, and Sara found that WNA has "political and financial motives." This is consistent with Kolsto's (2001) study in which the power company was found to be unreliable because it had a financial interest in the issue.

The Dale Academy students were given four choices for the data source—two blogs (the Ecocentric blog associated with *Time* magazine and the *Guardian*-associated blog), the Greenpeace website, and the World Nuclear Association website. Interestingly, the students did not classify the *Guardian*'s website as a blog, while *Time* magazine's website was categorized as a blog—possibly because it has the word "blog" in the title. Some groups dismissed the Ecocentric blog because the blog format was seen as unreliable—Adam, Andrew, Zu-Zhang, and Mario stated that blogs are generally not edited and proofread for content. More bluntly, David, Jared, Luca, and Clint did not trust websites that were not .org or .uk. These concerns about online content are also echoed in the Levinson et al. (2012) study in which two groups were questioning the reliability of information online. For example, a math teacher (Ella) finds a source questionable because it could be from "any old website…could be one person" (p. 221). The students in my study, however, seemed to have a heuristic which they used to differentiate between online sources (e.g., blogs are not reliable) rather than being sceptical towards all information found online.

Levinson et al. (2012) found that students "recognize the role of trust and authority in giving meaning to the data" (p. 224). Consistent with this, two groups in my study found the World Nuclear Association (WNA) most reliable because they are the authority on the issue and they are involved in the production of information: Blair, Federico, Robert, and Cai stated that WNA was "an organization measuring those things"; Jordan, Alex, and Aaron echoed this sentiment by saying that WNA was "the source that records the actual data." This trust in people (authoritative bodies) that are involved in direct observation and measurement is consistent with the findings in Levinson et al. (2012), in which they reported that a participant (Linda) finds the spine doctor's recommendation reliable because he "knows more about it than the other people, and he's seen more of these people" (p. 224).

Gal (2004a, 2004b) lists the questions of validity of data—arguments and data that support claims made within the data source—as an important component of the critical question which assist in the critical evaluation of data and the claims made about the data. As mentioned above, there is a solid body of evidence that the students considered reliability but there is virtually no evidence that the students questioned claims made within the specific data source (the four websites). This is consistent with Kolsto's (2006) finding that pupils evaluated sources of knowledge more than they evaluated the content of the statement.

However, the students did consider the presentation of data within the document. Many students (e.g., Adam from Dale Academy) claimed that they chose the *Guardian*'s website because of the data presentation: the accidents were given a numerical ranking in terms of severity with the addition of the detailed description of the incidents as well as the use of colour to designate severity.

Determining and Interpreting Probability

Probability Estimates. Determination of probability was presented to the students as the calculation of empirical probability. The estimation depended on the group's decision on the data source from the previous step. Findings showed that students used various ways of calculating empirical probability. However, even the groups who used the same data source (e.g., *Guardian*) differed on the data that was included in the calculation. For example, David, Jared, Luca, and Clint used the cut-off value of 5 on the INES scale, bringing the number of nuclear accidents to 6, whereas Daniel's group also used a cut-off value, but 4 instead of 5. These values depended on the students' evaluation and judgment about the "severity" of accidents.

This can be explained by the Levison's et al. (2012) pedagogic model of risk since the estimation of probability was influenced by the students' beliefs on what constitutes a serious accident. However, it also shows that the decisions about impact (interpretation of INES scale) are not separate from the decisions about probability, suggesting that the estimation of probability is also dependent on the estimation of impact, which is not explicitly stated in the Levinson et al. pedagogical model.

Representation of Probability. In the Levinson et al. (2012) study, the teachers' representation of probability (fractions versus natural frequencies) depended on the context. The findings in my study confirm the context-dependency of representations. Some of the groups in my study moved between different representations as they were making sense of the data, and also to defend their claims. For example, Federico, Blair, Cai, and Robert calculated that the probability of a nuclear accident was 0.00016 accidents per day, but they also represented the answer as 3 accidents in 18747 days, or alternatively 1 accident in 6249 days. However, later, when the groups were asked whether this accident showed that nuclear power plants were safe, they said they were. They supported their claim by saying that there was one accident in 17, yet again changing the representation—in this case the representation was used to make the number more approachable, as the interval of 6248 days may not be as tangible as 17 years. This shows that the interpretation of the estimate of risk as being small (on the far right of the Levinson et al. model, Figure 2) has influenced the representation. This is something that seems not to be explicit in the Levinson et al. model, since in the Levinson et al. model the representation of probability influences the estimation of probability, whereas the converse statement is neither documented nor discussed.

Lina, Louvie, Connie, Karl, and Andy also went from the decimal representation (0.55) to one involving frequencies, saying that there is 1 accident every 2 years. The preference for natural numbers has been widely documented by Zhu and Gigerenzer (2006), whose research shows that individuals' probabilistic reasoning is improved if natural frequencies are used instead of probabilistic estimates (fractions and decimals).

Some of the groups in both schools showed elements of being on the highest level of Watson and Callingham's (2003) construct (critical reasoning about rate). However, some students showed a lack of understanding about rational numbers, for instance, incorrect use of percentages amongst one of the St. Hubertus groups. Both groups also made mechanical mistakes in the calculation and communication of the results; there was no mechanism for checking the work amongst the groups.

There was also a lack of contextual understanding of rates. For example, in both cases, there were instances of students calculating the ratio of accidents per nuclear reactor by using the current number of reactors and the total number of accidents (since the 1950s), whereas the more consistent way would be to use the total number of reactors, past and present. Only one group used the notion of reactor years (days) which is the most common way to express operation of nuclear power plants amongst the experts in the field of risk assessment.

Interpretation and Decision Making Based on Probability. Once the students produced the value for the probability (the probability estimate), they were asked whether this indicated that the probability was small, and then, consequently, they were asked to decide on the safety of nuclear power plants. The findings showed that all of the Dale Academy students found the numbers to be small, whereas at St. Hubertus three groups found the numbers to be large and three found them to be small. Interestingly enough, Adam, Andrew, Zu-Zhang, and Mario found 0.076 accidents per reactor to be a small number, whereas Chloe, Mina, Larissa, Dan, and Sara found the same value to be large, arguing that 7.6 accidents per 100 reactors is not acceptable. When asked about the value, Adam said that they also took their personal views into consideration. This confirms Pratt and Levinson's claim that the interpretation depends on students' beliefs and also supports the Levinson et al. (2012) pedagogical model.

The relative frequencies students used for estimating probability had units (accidents/year). (My instruction at Dale Academy did not specifically suggest that students use units, whereas St. Hubertus students were instructed to use them.) The use of units may have given additional context and meaning to the probabilities.

Determining and Interpreting Impact

Qualitative Representation of Impact. There is evidence that students possess pre-existing informal (intuitive) knowledge of impact. From the pedagogic view, this is very encouraging because, in many other domains (such as assessment of probability), there is strong evidence that individuals' intuitions are often erroneous (see, for example, Kahneman, Slovic, & Tversky, 1982). As it was seen from the study, this informal knowledge has a potential to be used in the instruction. The initial assessment in both classrooms shows that students use different language to talk about impact (e.g., “massive” and “dangerous”). Other words used to express impact include: “big,” “astronomical,” and the students even used the phrase “a barren landscape” to visualize the impact of the Chernobyl accident. This corresponds to what Pratt et al. (2011) label as “fuzzy qualitative descriptors,” which students in his study used for the purpose of a rough quantification of impact ("serious," "massive," "bad," "fine," "big") (p. 339).

This is also consistent with the Levinson et al. (2012) study in which the teachers used phrases such as "impact on her life," "pain threshold," and "prohibitively dangerous option" to describe impact. The authors state that there was no opportunity for teachers to quantify impact. They also suggest that a meaningful quantification of impact and probability can only be done in an inquiry-based approach where the students can apply their values, representations, and experiences. This is the reason why the students in my study were given the impact statistic, and why I operationalized impact in terms of accidents and fatalities. Reasoning based on the magnitude of impact leads to reasoning about rational numbers—proportions, rates, and reciprocal values—confirming Watson's thesis that proportional thinking should be in the foreground of research about data. The findings showed that the students were making use of equivalent fractions, rates, and percentages. Sometimes they were correct (Blair's group, described above) but sometimes the use of rational numbers was incorrect. For example, some students were incorrectly using percentages.

Some students did not recognize that the statements were equivalent: 1.72 was the same as 31/18 which is the same as saying that the ratio was 1/18 to 1/31. However, it could be that the students thought that those statements, although mathematically equivalent, conveyed different contextual information. This stresses the differences between mathematical reasoning and quantitative reasoning in context (Mayes, Peterson, & Bonilla, 2012). The study as presented presents the case for the quantitative reasoning in context.

Another quantitative concept that the students were having difficulties with was the concept of reciprocal values. For example, St. Hubertus students understood that 31 fatalities/accident was a greater risk than 18 fatalities/accident. However, they did not understand that 1/31 accidents/fatality was the greater risk than 1/18 accidents/fatality.

Mathematical instruction underplays the importance of units. In the quantitative reasoning in context, however, the units are very important. My study shows that units were seldom used by Dale Academy students, while they were more often used by St. Hubertus students, which may be a consequence of them having been explicitly instructed to use units.

Coordination Between Likelihood and Impact. Unlike the studies of Levinson et al. (2012) and Pratt et al. (2011), the students in my study attempted to quantify both probability and impact. However, once quantified, there was a question of what to do with obtained numerical values. The students used various techniques to coordinate probability and impact—Dale students were instructed to multiply the two numbers, whereas the students at St. Hubertus used the graphing method (as suggested by Pratt et al.). Reflection on the Dale Academy students' use of the utility model (multiplication formula) confirms the observation made by Pratt et al. (2011) that, by reducing decision making to the product formula may have compromised the complexity, rendering the "decision-making process … irrelevant and meaningless" (p. 442). The Dale students' result of the product formula simply confirmed their overwhelming consensus that nuclear power is safe compared to other means of producing energy. The one group that disagreed decided not to use numerical information at all; instead, it used the qualitative descriptor of the impact of nuclear power plant accident as "astronomical."

The St. Hubertus students' use of the graphing method enabled them to visualize the relationship between likelihood and impact, and gave them more diverse ways of coordination. This is consistent with Pratt's call for the offering of other methods of quantification (p. 442). Mathematically speaking, the graphing method can be made equivalent to the product since the area of the rectangle defined by the point and the axes equals the product of the likelihood and impact.

However, in terms of the QRC framework, the graphing gives students an opportunity to locate the risk and then make risk management decisions about how to interpret it. For example, the area of the rectangle corresponding to nuclear power plant accidents—small probability, large impact—can be interpreted as medium risk; however, the students can consider contingency planning in terms of the accident. The graphing method does not necessarily make students make more objective decisions (the St. Hubertus students were still in line with their previous beliefs), however, it preserves the complexity by giving them the opportunity to use the language of risk. Also, it enables students to draw on ethical and other value-laden resources.

Some students did use the quantitative information to locate likelihood and impact on the axes, whereas some students used qualitative information. Another group also used the graph as an opportunity to ask the question about correlation between likelihood and impact—a group hypothesized that the larger the impact the smaller the probability, again leaving room for the richer interpretation.

Another characteristic of the interplay between likelihood and impact is that it does not happen only in the "final stages." The students considered likelihood and impact at various stages. For example, some Dale Academy students considered impact when determining the probability of a nuclear power accident by considering the cut-off value. Similarly, when asked whether 7% probability was large or small, Adam stated that it is small based on impact because the negative impact is smaller than the benefits.

The Role of Context and the Content Knowledge

The role of context. Pratt et al. (2011) discuss whether context may impede students' understanding of risk, drawing on examples from previous studies in which it was shown that context may be detrimental to mathematical understanding. The authors conclude that the understanding of risk is closer to statistics than to mathematics and that the context is crucial. If we strip away context and reduce the task to the mathematical coordination between likelihood and impact, we can see that the meaning is lost. The numbers have to be viewed in context.

The role of content knowledge. Levinson et al. (2012) state that "there were many opportunities in the microworld for the teachers…to make use of relevant scientific knowledge in helping to evaluate risk, but none chose to do so, reflecting other accounts, where scientific knowledge and information are either marginal or irrelevant to lay decision making" (p. 228). In my study, I found otherwise—the students' content knowledge interacted with their knowledge of risk, each at times elucidating the other. The students did seize the opportunity to use content knowledge, as the students seemed to understand that determination of risk and its various components depended on the context. For example, the knowledge about impact depended on students' being able to recall the information about nuclear power plants. However, not only did the content knowledge influence the knowledge about risk—the converse was also true. For example, in order to determine the impact of the nuclear power plant accident, the students at St. Hubertus drew on their knowledge of the content; however, as they were trying to understand impact, they were also making sense of the society.

How the question was formed was also an important part of the context in which students reasoned about risk. If the question had been framed differently, that would have influenced the study. The way I posed the question was whether nuclear power plants were safe relative to other energy sources. Also, the students understood the question to be whether "we" should have nuclear power plants. Some students made personal connections while others did not. This is consistent with the Pratt et al. (2011) study in which the students said that they would react differently depending on whether they were making decisions about somebody else or about themselves.

The role of feelings, beliefs and values. The affective factor is very important in individuals' risk based reasoning. Slovic (2010) discusses the dread factor that creates mental images about the hazards of interest (e.g., nuclear power plant accident). We can infer the feeling of dread in some imagery expressed by the St. Hubertus students, for example, a student talking about the impact of nuclear power plants as "a barren landscape." Similarly, Gregory et al. (2012) have shown that beliefs and values have to be an integral part of risk assessment and that the choice of data and the presentation of data depend on values. This can be seen in my study when Christine's group encountered the table of fatalities and argued about whether it was valid to only consider the accidents that resulted in five immediate fatalities. One of the students had a stern belief that "every death should count," and the other one was more pragmatic. Finally, the student drew on her personal experience, saying that it would matter to her if she was the person or if she knew the person. The students did not draw as much on personal experience as did the students in the Pratt et al. (2011) study. The reason is that the question was framed in terms of the logical statements: Are nuclear power plants safe? This can be compared to the decision statement: Should we have nuclear power plants, or more specifically, should we build more power plants in a certain area? Students did draw on their personal experiences, however. Particularly, one of the Dale Academy students was in the region (Hong Kong) when the Fukushima accident happened and he drew on this experience when making a decision about the safety of nuclear energy. In addition, another student at St. Hubertus stated that she would not like to live next to the nuclear power plant. However, because of how the question was construed, the students did not draw too much on personal experiences. Some students did show empathy (praying for Japan).

The pattern in both case studies was that the students did not seem to shift their beliefs about nuclear power plants (except in the case of one student). There were instances in which the exposition of quantitative data did cause students to question their beliefs. For example, some students were very surprised to find out that the fatalities for coal power plants were higher than those for nuclear plants. This is consistent with Kolsto's (2006) claim that students should be confronted with diverse information and viewpoints. However, students tended to include auxiliary information in order to "salvage" their beliefs.

Implications for Further Research

Some of the existing research suggests gender, race, and socio-economic differences play a role in perceived risk (Finucane et al., 2010), suggesting that white males tend to downplay risks compared to non-white males, white females, and non-white females. Although there is evidence that there are differences in the estimation of risk between the two classes, I am reluctant to make any conclusions based on gender, race, or socio-economic status because not enough data was collected about students' individual histories, identities, and beliefs about the world. There is a potential, however, to conduct research that would explore the interplay between different aspects of students' identity, beliefs, and understanding of risk in the classroom setting.

There also seemed to be gender differences in rhetorical style. Thirteen St. Hubertus students (more than half) expressed uncertainty of their knowledge about the accident (using phrases such as "I think" or "I don't know"), whereas none of the Dale Academy students did so. The expression of uncertainty did not necessarily match with the lack of knowledge. For example, Hiroko, who received the score of 2.5, prefaced her account of the events with, "I don't know much about what happened, but…” St. Hubertus students also expressed their emotions in their description, whereas only one student at Dale Academy did so—the student who had been in Hong Kong had been worried about the accident.

The question of the risk of nuclear power plant accident was framed in terms of a judgment about safety. Another way to frame the question would be in terms of the risk decision, such as whether there should be nuclear power plants. More specifically, we could frame a socially authentic inquiry-based learning instruction placed in the students’ own community. For example, students could assess the risks of having a uranium pellet factory in their neighbourhood. Framing the question in terms of decision rather than opinion may give more weight to the project, and it could create an opportunity for students to reflect on the relationship between their values, mathematical and statistical thinking, and the content knowledge. To echo Monteiro and Ainley (2003), the pedagogical setting of the task would resemble the inquiry setting.

The study also showed an importance of representations in reasoning about risk that goes beyond verbal and formal-mathematical representations. For example, my findings show that students were very often using gestures in order to elucidate their arguments. Another potential direction would be to research students’ use of gestures and locate it within other research on students’ gestures in mathematics. The students were also making sense of the documents containing data, very often choosing data sources that presented data in visually accessible ways. Another potential research direction would be to study visualization of risk data most accessible to students in the classroom setting. As a starting point, the research could apply existing research in the field of data visualization, such as that of Tufte (1997).

Conclusion

This article documents the complexity of the concept of risk and decision making based on risk. It also suggests how risk can be taught in the mathematics classroom. The study contributes to educational research by shedding light on the teaching and learning of risk in the mathematics classroom, whereas there is a lack of research in this area (Pratt et al., 2011). Another major contribution of the research is the identification of the understanding of rational numbers as being crucial to understanding risk.

An important lesson to take from this research is that decision making about risk is an interplay between quantitative reasoning, experiences, values, beliefs, and content knowledge. Restricting the instruction to any of these single components without meaningful consideration of other components will trivialize and reduce the effectiveness of the teaching. Risk is all around us and the pedagogy of risk should play an important role in mathematics education.

References

Aikenhead, G. S. (1985). Collective decision making in the social context of science. *Science Education, 69*(4), 453–475.

Akatugba, A.H. & Wallace, J. (1999). Mathematical dimensions of students’ use of proportional reasoning in high school physics. *School Science and Mathematics*, *99*(1), 31–40.

Beck, U. (2009). *World at risk*. Cambridge, MA: Polity.

Bybee, R., et al. (2006). *The BCSC 5E Instructional Model: Origins and effectiveness*. Retrieved from http://sharepoint.snoqualmie.k12.wa.us/mshs/ramseyerd/Science%20Inquiry%201%2020112012/What%20is%20Inquiry%20Sciecne%20(long%20version).pdf

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1–73.

Douglas, M. (1992). *Risk and blame: Essays in cultural theory*. London: Routledge.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, *34*(6), 1340–1357.

Fischhoff, B., Slovic P., Lichtenstein S., Read S. S., & Combs B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, 9, 127–152.

Finucane, M., Slovic, P. Mertz, C. K., Flynn, J., & Satterfield, T. (2010). Gender, race and perceived risk: The 'white-male' effect. In Slovic, P. (Ed.), *The feeling of risk: New perspectives on risk perception* (pp. 21–36). New York: Routledge.

Gal, I. (2004a). Statistical literacy: Meanings, components, responsibilities. In Ben-Zvi, D., & Garfield, J. (Eds.), *Challenges of developing statistical literacy, reasoning, and thinking* (pp. 47–78). Dordrecht: Kluwer Academic Publishers.

Gal, I. (2004b). A brief look at statistical literacy. *Math Practitioner*, *10*(2), 4–8.

Gal, I. (2005). Towards "probability literacy" for all citizens: building blocks and instructional dilemmas. In Jones, G. A. (Ed.), *Exploring probability in school*: *Challenges for teaching and learning* (pp. 39–63). New York: Springer.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In Wright, G., & Ayton, P. (Eds.), *Subjective Probability* (pp. 129–161). Oxford, England: John Wiley and Sons.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., & Ohlson, D. (2012). *Structured decision making: A practical guide to environmental management choices*. Chichester, UK: Wiley-Blackwell.

Greenpeace(2010). Calendar of nuclear events. Retrieved from: http://pec.putney.net/issue_detail.php?ID=18

Hansson, S. O. (2009). Risk and safety in technology. In: Meijers, A. (Ed). *Handbook of the philosophy of science: Philosophy of technology and engineering sciences, vol.* 9 (pp. 1069–1102). Oxford: Elsevier.

Harrell, B. (2011, March). Nuclear safety: US 'near misses' in 2010. Retrieved from: http://ecocentric.blogs.time.com/2011/03/17/nuclear-safety-u-s-near-misses-in-2011

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107.

Irwin, A., & Michael, M. (2003). *Science, social theory, and public knowledge*. Maidenhead, England: Open University Press.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. Educational Studies in Mathematics, 32, 101–125.

Kahan, D. M., Slovic, P., Braman, D., & Gastil, J. (2010). Fear of democracy: A cultural evaluation of Sunstein on risk. In Slovic, P. (Ed.), *The feeling of risk: New perspectives on risk perception* (pp. 183–213). New York: Routledge.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and Biases: The Psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kilinc, A., Boyes, E., & Stanisstreet, M. (2013). Exploring students' ideas about risks and benefits of nuclear power using risk perception theories. *Journal of Science Education and Technology*, 22, 252–266.

Kolsto, S. D. (2001). "To trust or not to trust…": Pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education*, 23, 877–901.

Kolsto, S. D. (2006). Patterns in students' argumentation confronted with a risk-focused socio-scientific issue. *International Journal of Science Education*, *28*(14), 1689–1716.

Lamon, S.J. (2007). Rational numbers and proportional reasoning. In Lester, F. K. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 629–667). Charlotte, NC: Information Age Publishing.

Langrall, C. W., & Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic thinking. In Jones, G. A. (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 95–119). New York: Springer.

Latour, B. (1993). *We have never been modern*. Cambridge, MA: Harvard University Press.

Latour, B. (2004). *Politics of nature: how to bring sciences into democracy*. Cambridge, MA: Harvard University Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. New York: Oxford University Press.

Levinson, R., & Rodd, M. (2009). Pre-service teachers' conceptions of risk: a pilot study. *Revista de Estudos Universitários*, *35*(2), 85–99.

Levinson, R., Kent, P., Pratt, D., Kapadia, R., & Yogui, C. (2011). Developing a pedagogy of risk in socio-scientific issues. *Journal of Biological Education*, *45*(3), 136–142.

Levinson, R., Kent, P., Pratt, D., Kapadia, R., and Yogui, C. (2012). Risk-based decision making in a scientific issue: A study of teachers discussing a dilemma through a microworld. *Science Education*, *96*(2), 212–233.

Lidskog, R., & Sundqvist, G. (2013). Sociology of Risk. In Roeser, S., Hillerbrand, R.,Sandin, P., Peterson, M. (Eds.), *Essentials of Risk Theory* (pp. 75–105). New York: Springer.

Mayes, R., Peterson, F., & Bonilla, R. (2012). Quantitative reasoning: Current state of understanding. In Mayes, R., & Hatfield, L. (Eds.), *WISDOMe: Quantitative reasoning and mathematical modeling: A driver for STEM integrated education and teaching in context* (pp. 7–38). Laramie, WY: University of Wyoming.

Moller, N. (2013). The concepts of risk and safety. In Roeser, S., Hillerbrand, R., Sandin, P., Peterson, M. (Eds.), *Essentials of Risk Theory* (pp. 75–105). New York: Springer.

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education*, *2*(3), 188–207.

Moore, D.S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Nilsson, P. (2007). Different ways in which students handle chance encounters in the explorative settings of a dice game. *Educational Studies in Mathematics*, *66*, 273–292.

Ontario Ministry of Education. (2005). *The Ontario Curriculum grades nine and ten: Mathematics (revised)*. Retrieved from http://www.edu.gov.on.ca/eng/curriculum/secondary/math910curr.pdf

Ontario Ministry of Education. (2011). *Environmental education: Scope and sequence of expectations*. Retrieved from http://www.edu.gov.on.ca/eng/curriculum/secondary/environ9to12curr.pdf

Polaki, M. V., Lefoka, P. J., & Jones, G. A. (2000). Developing a cognitive framework for describing and predicting Basotho students' probabilistic thinking. *Boleswa Educational Research Journal*, *17*, 1–21.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning, 13*(4), 322–345.

Pratt, D., & Yogui, C. (2010, August). *A constructionist approach to a contested area of knowledge*. Presented at the Constructionism Conference, Paris.

Rothman, R. L., Montori, V. M., Cherrington, A., & Pigone, M. P. (2008). Perspective: The role of numeracy in healthcare. *Journal of Health Communication, 13,* 583–595.

Science and Environmental Health Network. (1998). *The Wingspread consensus statement on the precautionary principle*. Retrieved from http://www.sehn.org/wing.html

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 463–489). New York: Cambridge University Press.

Slovic, P. (2010). *The feeling of risk: New perspectives on risk perception*. New York: Routledge.

Slovic P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2010). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. In Slovic, P. (Ed.), *The feeling of risk: New perspectives on risk perception* (pp. 21–36). New York: Routledge.

Tufte, E. (1997). *Visual explanations: Images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press.

Watson, J. M. (1997). Assessing statistical thinking using the media. In Gal, I., & Garfield, J. B. (Eds), *The assessment challenge in statistics education* (pp. 107–121). Amsterdam: The International Statistical Institute.

Watson, J. M., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.

Watson, J. M., & Shaughnessy, J. M. (2004). Proportional reasoning: Lessons from research in data and chance. *Mathematics Teaching in the Middle School, 10*, 104–109.

Wiener, J.B. (2002). Precaution in a multirisk world. In Paustenbach, D. (Ed.), *Human and ecological risk assessment: Theory and practice* (pp. 1509–1531). New York: John Wiley and Sons.

World Nuclear Association. Safety of nuclear power reactors. Retrieved from: http://www.world-nuclear.org/info/Safety-and-Security/Safety-of-Plants/Safety-of-Nuclear-Power-Reactors/

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition, 98*, 287–308.

Zieffler, A., Garfield, J., & delMas, R. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40–58.

Risk Intuitions and Perceptions: A Case Study of Four Year 13 (Grade 12) Students

Stephanie Budgett
The University of Auckland, New Zealand
Lorraine O'Carroll
The University of Auckland, New Zealand
Maxine Pfannkuch
The University of Auckland, New Zealand

Abstract: In the New Zealand school statistics curriculum, year 12 students (aged 16-17) are required to solve problems that involve interpreting risk and relative risk within a range of meaningful contexts. In a small exploratory study we investigate the risk conceptions of four year 13 students who performed at the excellence level in their year 12 externally-assessed examination on this topic. Through questionnaires and interviews we investigate the ways in which these students perceive and express risks associated with a variety of everyday activities and also how they compare the risks of several adverse outcomes. We also explore the strategies they use when confronted with varied representations of risk such as visual, verbal and numerical. We will report on insights gained about these students' reasoning with different risk representations, on how they interpret, evaluate and communicate risk.

Keywords: secondary school students; heuristics; risk perceptions; risk representations; risk interpretations; relative risk.

Introduction

Statistical information is prolific in the media, yet many of us are misled, or have difficulty in interpreting and challenging statements made. In order to be an educated citizen in the twenty-first century, "(S)tatistical literacy is becoming fundamental for living in a full democracy" (Biggeri & Zuliani, 1999, p. 2). An important facet of statistical literacy is the ability to reason and to make judgments when an element of uncertainty is involved. We tend to rely on our gut feelings, combined with personal experiences or beliefs in order to make such decisions, and often use shortcuts such as the availability and representative heuristics (Kahneman, 2011). However, this approach may fail us when circumstances require more of an analytic assessment. Examples of such situations include the choice of a particular medical plan, the decision to invest in stocks or shares, assessing tsunami risk and measuring the effects of changes to scheduling of transport systems. Although sometimes trivial, the consequences of such decisions can be significant. Thus an understanding of risk is important in order to be an educated citizen in the 21st century.

The topic of risk was introduced to the New Zealand school statistics curriculum in 2012. The result of this is that year 12 students (aged 16-17) are now required to solve problems that involve calculating and interpreting risk and relative risk within a range of meaningful contexts. In addition to concepts such as risk and relative risk, the topic involves probability distributions, relative frequencies, two-way tables and probability trees. Students need to be able to demonstrate knowledge of probability concepts and terms, and to communicate using appropriate representations. This small exploratory study, involving four students, is a first attempt to explore how these students now perceive and express risks associated with every day activities, how they compare risks of several adverse outcomes and what strategies they use when confronted with varied representations of risk such as verbal, visual and numerical.

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 330-346

Background Literature

What is Risk?

The Oxford English Dictionary definition of the word 'risk' is "*(Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance; a chance or situation involving such a possibility*". The everyday definition of risk will depend on the context in which it is being used. One common way of defining risk is as an uncertainty based on historical observations (Gigerenzer, 2002; Johnson, 2004) with other definitions including a hazard, or a potential adversity or threat that requires exposure and uncertainty (Slovic & Weber, 2002). As a concept, risk can refer to a probability or to a consequence or to the product of probability and consequence. For example, to answer the question "what is the risk of getting cancer if I eat processed meat frequently?" one might respond with a numerical value such as 0.01 or 1%. On the other hand, to answer the question "what is the risk of defaulting on my mortgage payments?" the response may be that my home will be repossessed by the bank. Gardner (2008) noted that risk can be perceived as the product of the probability of an event occurring and the value placed on its consequence. A study carried out by Sadique, Devlin, Edmunds, and Parkin (2013), explored decisions made by mothers who were given a hypothetical situation and asked whether or not they would vaccinate their child against a disease. The severity of the health effects associated with the disease was explained as was the severity of the adverse side-effects due to the vaccine, together with the probabilities of their occurrence. Their findings suggested that the information regarding consequence, in this case the severity of the health effects associated with the disease and the severity of the adverse side-effects due to the vaccination, influenced the mothers' decision to vaccinate or not, while "the probability of these events occurring was not a significant predictor" (p. 1). The very fact that the term risk encompasses multiple definitions and concepts results in many of us having trouble in both interpreting and communicating risk (Slovic & Weber, 2002).

Communication of Risk

It is a well-established fact that our perception of risk is influenced by the way in which that risk is presented (Gigerenzer & Edwards, 2003; Tversky & Kahneman, 1981; Utts, 2005). Since the provider of risk information is at liberty to choose how to convey the information, they are free to select the method that best serves their interest. An example might be a drug company quoting the benefit of their treatment in terms of relative risk instead of absolute risk in order to make results more compelling. Thus, owing to the many and varied ways in which risk can be communicated, we are easily manipulated. In a risk-literate society, citizens should have the ability to interpret risk in its many forms in order to make informed life-decisions (Gal, 2005). It is therefore desirable for learners to experience risk information in several formats, and to appreciate that the provider of information may present information in a manner that suits their own agenda, and not necessarily that of the consumer.

Verbal communication. Verbal representations allow for flexibility in risk communication and, as such, have the potential to capture consumers' emotions and intuitions in a way that numerical and graphical representations cannot. While this may enhance risk comprehension in some instances, there is also the potential for confusion and misunderstanding, particularly when the communicator of the risk information has an agenda to push. In particular, the manner in which information is framed, the *verbal frame*, can affect the way in which that information is perceived and acted upon. For example, when persuading patients to undertake a risky treatment option, evidence suggests that presenting probabilistic or risk information in a positive frame is more effective than presenting it in a negative frame (Edwards, Elwyn, Covey, Mathews, & Pill, 2001). However, when considering how to communicate the risks associated with taking part in a screening programme it appears that loss framing, where the potential losses from refusing to participate in the programme are described, has a greater impact on participation rates than gain framing, where potential gains from agreeing to participate in the screening programme are described (Gigerenzer & Edwards, 2003).

Visual communication. Visual displays of information, whether static or animated, have the potential to capture attention, summarise data effectively and to reveal potentially undetected patterns such as part-to-whole relationships (Eppler & Aeschimann, 2009). However, they are not immune to manipulation by risk information providers. When considering the design of a visual risk representation, it is important to consider the audience and to ensure that important details are highlighted, and superfluous details are avoided. Effective visual representations need to be designed carefully so that misinterpretation is minimised. Often, simpler graphs are more effective than more complicated graphs (Zikmund-Fisher, Fagerlin, & Ubel, 2010).

Numerical communication. Since a common way of representing and communicating risk is in numerical format, consumers of such information require adequate numeracy skills (Fagerlin, Zikmund-Fisher, Ubel, Janlovic, Derry, & Smith, 2007). However, even with sufficient proficiency, the way in which risks are expressed mathematically has an influence on their interpretation (Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). Evidence suggests that expressing risks with natural frequencies lends itself to easier interpretation than proportions, probabilities or percentages since this is how statistical information has been presented historically, and the human mind is well-adapted to information in this form (Gigerenzer, Gaissmeier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Martignon, 2014; Peters, Hibbard, Slovic, & Dieckmann, 2007).

Perception of Risk, Heuristics and Biases

People tend to perceive and react to risk in one of two ways; either as *risk as analysis*, or as *risk as feelings* (Slovic & Peters, 2006). While risk as analysis refers to a logical and rational treatment of risk information, risk as feelings refers to a reflexive reaction which may be based on no more than a hunch or a gut feeling. When judging or interpreting risk as feelings, we are susceptible to appealing shortcuts or heuristics which, although reliable in some instances, may also lead us astray. The availability heuristic is a shortcut we use that allows us to make a judgment based on how easy it is to recall a particular situation (Kahneman, 2011). For example, we tend to think that deaths due to causes such as murder, airplane accidents and shark attacks are more common than deaths due to less memorable causes such as flu, heart disease and diabetes due in part to the well-documented tendency for the media to devote more column-inches to relatively rare events than to conditions or behaviours that carry higher risks (Aronson, 2006; Harrabin, Coote, & Allen, 2003; Thirlaway & Heggs, 2010). This phenomenon results in the availability heuristic leading us astray. However, the availability heuristic may also serve us well. For example, if we find ourselves in a problematic position, we may recall unpleasant consequences occurring in a similar situation, and act to better protect ourselves. Another common shortcut is the control heuristic. Thompson, Armstrong, and Thomas (1998) suggest that people often adopt the control heuristic when assessing their personal influence over the occurrence of an outcome, even in chance situations. The control heuristic is comprised of both the intention to achieve an outcome and the perceived connection between one's action and the outcome. When we perceive a connection between action and outcome, the notion of personal control will be high. As with the availability heuristic, the control heuristic can lead to accurate judgments in situations such as those where we do in fact have control. However, in situations where we have little or no control, the control heuristic can give rise to erroneous decisions.

There are also several biases to which we are susceptible when receiving and interpreting risk information. According to Weinstein (1989, p. 1232), the optimism bias has the result that "people regard themselves as more likely than others to experience financial success, career advancement, and long life". This bias results in the belief that one's own risk is lower than the risk faced by other individuals sharing the same behaviour. Other examples of the optimism bias in action include judging our driving skills as higher than average, and our chance of being involved in traffic accidents being lower than for others (Svenson, Fischhoff, & MacGregor, 1985). The bias of risk denial arises when evidence is considered as reliable and informative only if it is consistent with one's beliefs, otherwise it is dismissed (Everitt, 1999). Another bias, the anchoring bias, relates to our inclination to rely heavily on one, often irrelevant, piece of information when making judgments, often evaluating new information against this anchor (Anderson & Iltis, 2008; Kahneman, 2011).

How Risk Should Be Taught

It seems indisputable that an understanding of risk is crucial in order for us to participate in the modern world. Thus, "training young students in the perception of risk has become fundamental in modern society", (Martignon, 2014, p. 157). According to Gal (2005), risk literacy is a branch of probability literacy that is closely related to statistical literacy. He describes five main knowledge classes and some dispositions which he proposes form the basic foundation of probability literacy. The knowledge elements comprise (1) big ideas (such as variation, randomness, and independence), (2) figuring probabilities, (3) language, (4) context, and (5) critical questions. Dispositional elements include critical stance, beliefs and attitudes, and personal sentiments such as risk aversion. Probability instruction in school has a tendency to be based on probability axioms and mathematical calculations (Greer & Mukhopadhyay, 2005; Jones, 2005). However, for the vast majority of us, the situations that necessitate us to draw on probability knowledge will be those requiring judgments or interpretations, not calculations. Gal (2005) has proposed some critical questions that consumers should use when interpreting statistical and probability claims. Based on these, and on what others consider that consumers need to ask about probability, and risk in particular (Gigerenzer et al., 2007; Utts, 2005), the following critical risk questions are proposed (see Table 1).

Big idea	*Details*
The risk of what?	• Baseline information • Risk communication (verbal, visual, numerical format)
What is the frame?	• Time frame • Verbal frame
To whom does the risk apply?	• Does it apply to me?

Table 1. Proposed critical questions for risk literacy

Risk in the New Zealand Mathematics and Statistics Curriculum

In 2012 the New Zealand Mathematics and Statistics Curriculum introduced a National Certificate of Educational Achievement (NCEA) standard at Level 2 (penultimate school year for most students) which in part relates to interpreting risk and relative risk. This standard, *Apply probability methods in solving problems*, requires students to exhibit proficiency in selecting and using methods, demonstrate knowledge of probability concepts and terms and to communicate using appropriate representations. The problems that are used are set in either real-life or statistical contexts, with probability methods including risk, relative risk, probability distributions, relative frequencies, two-way tables and probability trees. A typical risk-based question would describe a study with a real-life context such as, for example, wishing to investigate the prevalence of a disease in individuals with differing characteristics. It may be of interest to establish whether one group of individuals is more susceptible to the disease than another group, or that treatment A is more effective in combatting the disease than treatment B. Information may be given in terms of frequency data, or in a two-way table. Students are expected to be able to calculate probabilities, conditional probabilities, absolute risks and relative risks and possibly to critique media statements based on findings from the study in question. Although interpreting risk formed a small part of this standard, it was of interest to determine to what extent students' experiences from the previous year remained with them as they were confronted with new risk situations and representations.

Methods

This study was exploratory in nature, with the purpose of developing a deeper awareness of the issues associated with learning about risk in the school curriculum. The participants in this study were four Year 13 (aged 16 and 17) students attending a single-sex girls' secondary school in New Zealand. The participants were selected randomly from a larger group of 20 students from the same school, all of whom achieved well. The second author collected the data which comprised questionnaire responses and interviews. The students first completed a questionnaire designed to elicit their understanding of risk in a variety of ways. Six of the nine questions in the questionnaire were from the work of Dargahi-Noubary & Growney (1998), one question was adapted from Iman (1994), one from Fagerlin et al. (2007), and one from a previous Year 12 national examination. The questions required the students to describe, calculate, compare and interpret risk in several contexts and, where relevant, were asked to explain their answers. Once the students had completed the questionnaires, two follow-up interviews were carried out with the students working in pairs. The purpose of the first interview was to gain more insight into student responses in the questionnaires, while the aim of the second interview was to explore their reasoning strategies when presented with varied, and sometimes unfamiliar, risk representations. The interviews were video-recorded and transcribed for analysis. A thematic qualitative data analysis (Braun & Clarke, 2006) was conducted on the student questionnaire responses and student interviews.

Analysis

Since similar themes were obtained from several questions and tasks given to the students, attention will focus on only five of the nine questions in the questionnaire, and on three of the four tasks that formed the second interview. In sections 4.1 to 4.3, we focus on Questions 1, 2, 3 and 6 of the questionnaire. In section 4.4 we present findings from the second interview where students were given four risk representation tasks. Three of the four tasks will be focused on, and at this stage Question 9 from the questionnaire will also be discussed.

Description/Definition of Risk

The first question (Dargahi-Noubary & Growney, 1998) required the students to state what they knew about the topic of risk based on their general knowledge, school work, or from what they had heard in the media. While not specifically asked to do so, three of the four students (Students A, B and C) provided their definition of risk. All of these students defined risk as the likelihood of an event occurring, or words to that effect. Two of these students stated that the event in question was undesirable, that is they associated risk with a negative outcome. Student B's response was "Risk is how likely something is to occur, usually in regards to a negative outcome", whereas Student C responded "Risk is how likely something bad is going to happen". Student A stated that there were "good risks and bad risks" while Student D answered the question by mentioning how she thought the media portrayed risk, and by stating that her school experiences taught her to consider "the consequences of our actions and risk which comes with the choices we make." This student is thinking on a broader level than the other three by reflecting on how risk is generally portrayed in society.

In Question 3 the students were asked to choose their preferred definition of the term risk from a list of four provided in Table 2.

Which of these definitions of risk do you prefer? Explain.

(a) A probability with a negative connotation

(b) A type of probability that has no theoretical model

(c) An uncertainty that has a numerical value

(d) Other – (explain)

Table 2. Risk definitions provided to students in Question 3 (Dargahi-Noubary & Growney, 1998, p. 45)

Given their responses to Question 1, it was not surprising that students B and C chose option (a). Despite her response to Question 1, Student A also selected option (a). Student D chose option (d) with the accompanying explanation that risk could refer to events with positive or negative outcomes. She also stated that the media tend to use the term risk when discussing negative outcomes.

We conjecture that there may be several reasons to explain why risk is associated with negative outcomes. One reason may be that the students cannot envisage the complete *sample space* of outcomes associated with an event under consideration. Confounding this issue is the fact that everyday *language* may interfere with risk perception so that even if students are able to imagine the complete sample space of outcomes, they may use different language to describe the likelihood of each outcome. In addition, the *availability heuristic* may be operating in such a way to influence the students' perception risk, a fact that Student D was aware of. As an example, consider testing the effectiveness of a new drug. A naïve consideration of all possible outcomes would be to state that the drug either worked, or did not work. However, not only would it be of interest to establish whether or not the drug worked, but also if the drug caused any side effects. Therefore, a complete list of outcomes may be: (1) the drug works and there are no side effects, (2) the drug works and there are side effects, (3) the drug doesn't work and there are no side effects, and (4) the drug doesn't work and there are side effects. When asked to assess the riskiness of the four outcomes, it may be that the *language* depends on the outcome. For example, the word 'chance' may be used to describe the likelihood of outcome (1) which describes a positive response, while the word 'risk' is used for the other outcomes which all include at least one negative response.

Ranking Risks

The second question, which was not part of the Dargahi-Noubary and Growney (1998) questionnaire, required the students to rank a list of ten causes of death according to their risk, with a rank of 1 being allocated to the event with the highest risk, and a rank of 10 to the event with the lowest risk. Eight of the causes of death were due to injury, either deliberate or accidental, with the remaining two (smoking and stroke) being disease-related. This question aimed to provide some insight into the students' perceptions of the comparative risks of several events and was based on similar questions adopted by others, for example Iman (1994) and Jones (2012). Students were then asked what background information they would need in order to estimate the risks associated with each event. The rankings from the students are provided in Table 3, with the mean rank being the average awarded across all four students. The final column in Table 3 gives the actual ranking of the deaths due to injury provided by the National Safety Council (National Safety Council, 2006), an organisation that collates data on such events for the United States, with the relative rankings of deaths due to smoking-related illnesses and strokes obtained from the Centers for Disease Control and Prevention website (www.cdc.gov).

Cause of death	Student				*Mean rank*	*NSC/CDC rank*
	A	*B*	*C*	*D*		
Falling	7	10	10	3	9	5
Firearm assault	9	1	6	9	7	6
Motorcycle accident	8	4	1	8	4	8
Motor vehicle accident	3	5	7	7	5	3
Pedestrian accident	2	8	3	4	3	7
Stroke	6	9	4	5	6	2
Smoking	4	7	9	6	8	1
Suicide	1	3	2	1	1	4
Natural forces	10	6	8	10	10	10
Drowning	5	2	5	2	2	9

Table 3. Students' and NSC/CDC rankings for Question 2

There were several interesting aspects to the student responses to Question 2. Firstly, students tended to rank *suicide* very high despite its actual ranking of 4. Secondly, while *smoking* and *stroke* ranked highest according to the data provided by the CDC, students placed both of these causes of death as low (8, 9) or moderately low (4-6). In addition, two students ranked *drowning* as their second highest cause of death, while the data provided by the NSC resulted in a rank of 9. Finally, two students ascribed the lowest possible rank of 10 to *falling*, while its relative ranking is 5. Possible explanations for these results include the use of the availability heuristic, a lack of awareness of critical questions, and the notion of conditioning on an event.

Availability Heuristic. The role that the *availability heuristic* plays is the possibility that the students are relying on media reports and subsequent conversations with their peer groups and/or family to make conclusions about the incidence of suicide, drowning and deaths from smoking or strokes. This study took place around the time when the media had reported on several high-profile celebrity suicides, thereby perhaps distorting the perception of the prevalence of suicide. In Student D's words, "as you grow older you are more familiar with death… from old age or suicide kind of because you hear a lot about it, but when it comes to being hit on a motorcycle you are like oh, I haven't heard that for ages". Cases of drowning also receive heightened media-attention, largely due to the fact that in New Zealand water-based activities such as fishing, swimming, surfing and sailing are popular leisure pursuits and the media therefore have a responsibility to raise an awareness of the dangers associated with these activities. Thus students may perceive drowning to be more common than it actually is. On the other hand, news reports about deaths due to strokes and smoking are few and far between. Given the tendency of the media to report prolifically on relatively rare events, this may have distorted the students' perceptions of the regularity with which some of these events occur.

Critical questions. It appears that these students did not consider the questions that should be asked, such as the critical questions outlined in Table 1, when ranking the events in Table 2. Taking *falling* as an example, it appears that the students may be considering that their own risk and that of their peers, of dying from a fall is very low and that that is the basis for the low ranking. They do not appear to consider the time frame over which the events are considered. If they had taken the time frame into account, then perhaps they would have recognised that on a population level the risk of dying from a fall is not insignificant due to the higher prevalence in the elderly. The same may be said for their comparatively low rankings for deaths from smoking and strokes, since they perhaps don't see these events as being relevant to themselves or to their peer groups. However, even though the students did not appear to reflect on the critical questions when ranking the events in Table 2, they appeared have an awareness of some of the questions when they were asked what information would be required in order to provide numerical estimates. Responses included a requirement to have information in the form of frequencies, for example Student B stated: "How often each of the events occurs" which suggests that she was considering the risks at a population level rather than at a local level. Student B also asked for information as to whom the risk applies, for example "How often does the person smoke?" and "Motor cycle accident – was the person wearing a helmet? What speed were they going?" thereby demonstrating an awareness of *sample space* and its relevance when required to make a judgment.

Conditioning on the event. Follow-up interviews with the students uncovered an interesting feature to one student's interpretation of this task. Student B explained that she allocated a rank of 3 to suicide and explained "I wasn't sure whether thinking … if you're going to commit suicide what is your chance of dying from it, it's pretty high. Like if you're already in the act…" and, for motor vehicle accidents "a lot of people survive car accidents, so it was probably lower than suicide" which she ranked as 5. Student B appears to be using conditioning in her reasoning by considering the risk of dying *given* each event is underway, and therefore ranking them according to the likelihood of surviving the event once it has happened. Despite this reasoning, her rankings for suicide and motor vehicle accidents were similar to the NSC data-based rankings. While assessing the relative rankings of the events in Table 1, Student B appears to be integrating knowledge about context and consequence as part of her reasoning strategy, giving her a sense of the size of risk that she might attribute to the events. Her belief is that once engaged in the act of suicide, the chance of survival is very low, while she believes there is a much higher chance of surviving a motor vehicle accident. Her contextual knowledge does not seem to include an awareness that there will be many unsuccessful suicide attempts which is not surprising given that such events are generally undocumented in the media and elsewhere.

Numbers Associated with Risk

Question 6 required the students to describe a risky activity in which they had been involved, and to estimate the amount of risk associated with that activity. They were then asked to consider how their friends might respond to the risk estimates for the same activity and to explain their reasoning (Dargahi-Noubary & Growney, 1998). The activity provided by Student A was that of going on an overnight hiking expedition in the bush, with negative outcomes comprising getting lost or injured, or suffering from hypothermia. She explained that there was the potential for such a hike to be considered a high-risk activity, but that the preparatory groundwork and the experience of the hike leaders meant that the actual risk was fairly low, giving a numerical value of 2%. In her response, Student A exhibits the personal control heuristic by considering that the preparatory actions of group would directly influence the possibility of a hazardous outcome. She then went on to explain that the risk estimates provided by friends who were involved in the same activity would depend on their individual experience levels and their perceptions of the competence of hike leaders.

Student C's example was of riding her unicycle to dance class. She felt that this was a high risk activity since a unicycle is difficult to control and can be very unstable on uneven ground. She gave a numerical value of 70% for her own risk of falling and explained that this was because of the "numerous things that could go wrong". She believed that her peers would have a much higher ("maybe 90%") chance of falling since they were less experienced. Again, there is evidence of the personal control heuristic in Student C's response although she is aware that, despite her level of experience, she is still vulnerable.

While Student A's risk estimate is very low (2%) compared with that of Student C (70%), it is worth mentioning that the negative outcomes associated with the hiking activity may well be perceived as being more serious than those associated with falling off a unicycle, particularly when protected by a cycling helmet. Therefore the severity of the *consequences* may be influencing the students' risk perceptions more than the actual likelihood of adverse events occurring.

Risk Representations

In order to gain some insight into how these four students reasoned about risk when presented with varied, and sometimes unfamiliar, risk representations, further tasks were given. In these tasks, students worked in pairs to understand and reason with risks presented both visually and numerically. The visual representations included both graphical formats with information being displayed as line graphs, and pictorial formats with information being displayed as icons. Numerical representations involved information being presented either in the form of probabilities, natural frequencies, or two-way tables.

Visual risk representations. Graphical information on the risk of invasive breast cancer diagnosis (Dupont & Plummer, 1996) was presented to the students (see Figure 1).

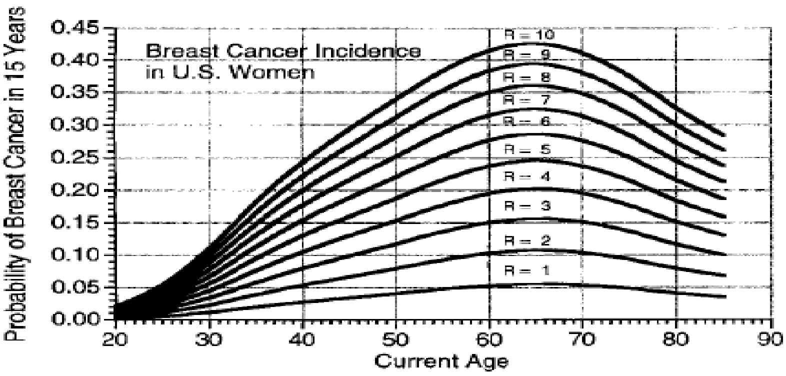


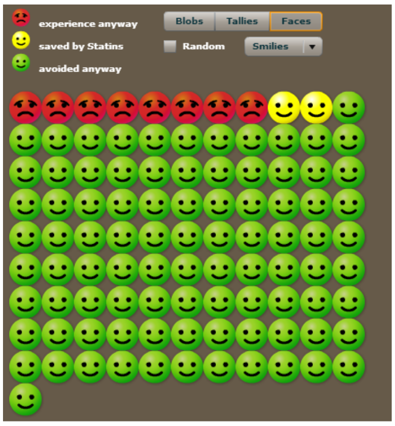
Figure 1. 15 year absolute risk of invasive breast cancer diagnosis by current age and relative risk (Dupont & Plummer, 1996, p. 2196)

This particular format would not have been familiar to the students. However, both pairs of students were able to read information from the graphs, and to think of some risk factors that might explain the different R-lines in the graph. Students A and B suggested that family history and lifestyle would be factors that may influence an individual's breast cancer risk, while Students C and D suggested exposure to UV light. When asked to estimate the probability of someone aged 60 with an R-line value of 4 being diagnosed with invasive breast cancer within a 15 year time period, all four students were able to interpret the graph correctly and provide a sensible answer. Students were then asked why all of the curves appeared to peak at the age of 65. Students A and B were able to make use of their background contextual knowledge with Student A stating: "…and maybe if you've had all those increased risk factors such as family history you'll have got it before then so if you've got all the way to 65 means you've got a good family history." When asked if a corresponding graph for women in New Zealand would look similar to Figure 1, which is for US women, Students A and B noted that the graph would look similar because "we live in very similar lifestyles and have a similar demographic to the US, we're a Western culture and if you went to Africa you'd probably get a different graph". On the other hand, Students C and D thought the graph would be different because in the US there are "different diagnostic techniques or not as many people in New Zealand have scans… so it won't be exactly the same percentage". While each pair of students answered this question differently from one another, their responses were sensible and indicated that they were all using their contextual knowledge about breast cancer, and integrating this into their reasoning.

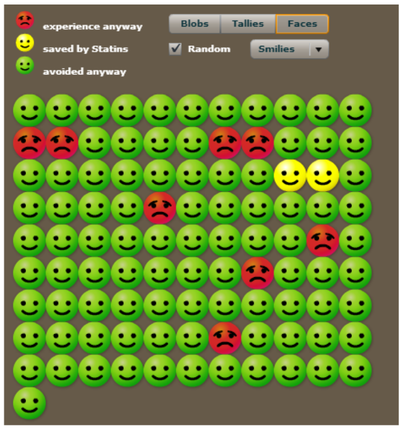
Two different situations, both incorporating pictorial risk information in the form of icons, were also provided (see Figure 2). The information related to the effects of statins (cholesterol-lowering drugs) on the incidence of heart attack/stroke (see www.understandinguncertainty.org website, Spiegelhalter). Scenario 1 used loss framing by focusing on the disadvantages of not taking statins while Scenario 2 used gain framing by describing the positive effects of taking statins. Within each of these scenarios the information was provided visually in two ways; either with grouped icons where like-icons are placed in groups as can be seen in Figures 2(a) and 2(c), or with randomly placed icons where like-icons are placed randomly as can be seen in Figures 2(b) and 2(d).

Scenario 1

Your chance of experiencing a heart attack or stroke in 10 years without Statins is in 100, which is reduced to___in 100 with Statins.



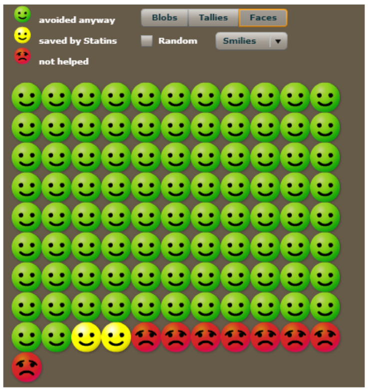
2(a)



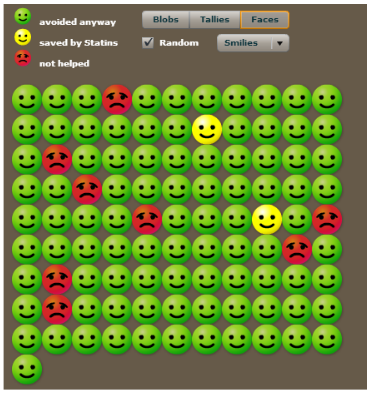
2(b)

Scenario 2

Your chance of avoiding a heart attack or stroke in 10 years without Statins is ___in 100, which is increased to__in 100 with Statins.



2(c)



2(d)

Figure 2. Pictorial representations for risks with loss framing (Scenario 1) and gain framing (Scenario 2) with grouped icons (a, c) or randomly placed icons (b, d) (http://understandinguncertainty.org/node/233)

The students understood the visual representations shown in Figure 2 and did not appear to have any trouble in answering the questions stated within each scenario. When asked for their opinions on the grouped icons versus randomly placed icons, Students A and B expressed a preference for grouped icons with Student A stating that they were "easier to understand conceptually whereas counting them wasn't really that hard but visually looking at it, it was a lot easier when they were grouped". Whether the icons were placed in groups or randomly did not appear to make a difference to Students C and D. The students were then asked if they had a preference for Scenario 1 or for Scenario 2. This questioning was aimed at discovering whether the verbal frame made a difference to the students' reasoning. Again, students C and D had no real preference while Students A and B mentioned that it depended on the numbers. Student A said: "It was interesting because they've both got a difference of two but with the 90s it seems like a smaller gap… so it seems more significant when you're talking about the 8 and the 10, even though they're both just two". Her statement suggests that the effects of the statins are more apparent within the loss frame of Scenario 1 because the difference between 8 and 10 is relatively larger than the corresponding difference between 90 and 92 when the situation is expressed within the gain frame of Scenario 2.

Numerical risk representations. In order to explore these students' reasoning when confronted with numerical risk information, a breast screening and diagnosis problem was presented in two ways; one probabilistically, the other with natural frequencies (Gigerenzer, 2002; see Table 4).

Version (a)	The probability that a woman of age 40 has breast cancer is about 1 per cent. If she has breast cancer, the probability she tests positive on a screening mammogram is 90 per cent. If she does not have breast cancer, the probability that she nevertheless tests positive is 9 per cent. What are the chances that a woman who tests positive actually has breast cancer?
Version (b)	Eight out of every 1000 women have breast cancer. Of these 8 women with breast cancer, 7 will have a positive mammogram. Of the remaining 992 women who don't have breast cancer, some 70 will have a positive mammogram. Imagine a sample of 1000 women who have a positive mammogram. How many of these women actually have breast cancer?

Table 4. Two versions of problem given to students (Gigerenzer, 2002)

Students A and B, who received version (a), immediately drew a probability tree and wrote the correct probabilities along the branches of the tree. As they started to multiply the probabilities along two sets of branches and add them together, they then decided that only the branch reflecting those with breast cancer who tested positive was needed. Student B stated "so isn't it just 90%?" and Student A agreed. Students C and D received version (b), but did not draw a probability tree. Instead, they tackled the problem verbally and concluded that 77 women had a positive mammogram, "and how many of these women actually have breast cancer? So isn't it just seven out of 77?". The students who were given the risk information in probabilistic format appeared to lose sight of the relevant sample space, convincing themselves incorrectly that it was women who have breast cancer. However, the students provided with numerical information gave the impression that they were able to retain information about the relevant sample space whilst discussing the problem, thereby arriving at the correct answer.

In the Question 9 of the questionnaire, which students answered individually, they were given a two-way table describing the outcomes (cured or not cured) of a group of individuals with a serious skin rash who either received treatment A or treatment B. They were asked to calculate the probability of being cured with treatment A, and then to critique a newspaper headline stating that a person with a serious skin rash is twice as likely to be cured if using treatment A as a person using treatment B. In order to answer this question, the students were required to calculate the probability of a person being cured with treatment B, and then to calculate the required relative risk. This question was very familiar to the students as it resembled typical questions asked in their probability standard of the previous year. All of the students answered this question well.

Discussion and Conclusion

This paper explores the strategies used by Year 13 students as they were asked to interpret, rank and reason with risks in a variety of contexts, and were confronted with a variety of risk representations. In this final section we discuss reasoning strategies and risk conceptions displayed by the four students and consider the contribution that the findings may have to existing research in this area. We then discuss the implications the findings have on teaching risk in the classroom. However, we will first comment on the limitations of the study.

Limitations of the Study

This exploratory study was very small, focusing as it did on only four high-achieving students. Within the constraints of the study, many potential areas for investigation were not possible. For example, it was not possible to explore the effects of cognitive limitations and biases and personal characteristics such as cultural background, which researchers have found to be contributory factors in interpreting and misinterpreting risk information (Anderson & Iltis, 2008; Sorensen, Gyrd-Hansen, Kristiansen, Nexoe, & Nielsen, 2008).

Potential Characterisations of Student Conceptions

While acknowledging the limitations of the study, the findings seem to suggest that certain potential characterisations of student conceptions are apparent (see Table 5). These conceptions, based on responses to all questions in the questionnaire and on all tasks in the second interview, will now be discussed further.

Conceptions	Description	Characterisation in study	Student Reasoning
Availability heuristic	Judgement is created on the availability or recall of information in the memory	• Likelihood of risks are given	Led to sensible use of own contextual knowledge to judge reasons for likelihoods
		• Likelihood of risks are estimated	Own contextual knowledge led to errors in judgements
Critical questions to ask when judging risk information	*Risk of what?*		
	• Baseline information	Size of risk depends on the baseline	Not taken into account when judging risk but mentioned when asked how to determine risk
	• Risk communication	Can be verbal, visual or numerical format	Appropriate reasoning seemed to depend on format
	What is the frame?		
	• Time frame	Risk of events depends on the time frame	Not considered when time frame was not given
	• Verbal frame	Information presented in positive/negative, harm/benefit frames influences estimation of risks	• Larger numbers influenced preference for positive frame • Preference for risk expressed as harm or benefit not evident
	To whom does the risk apply? Does it apply to me?	Judgements of risks are influenced by identification of relevant sample space	Relevant sample space not taken into account unless given
Reasoning numerically with risk	• Estimating risk	Sense of size of risk (e.g. 10%)	• Considered context and consequence of risk • Developing a sense of associating numbers appropriately with high and low risk • Personal control ideas tended to lead to a disregard of universal sample space

	• Calculating risk	• Natural frequencies • Proportions	Led to appropriate calculations Led to inappropriate calculations
	• Language of risk	Everyday language is different from probability language. Probability and risk calculations are the same	Some perceived risk as having a negative connotation
Sample space	• Universal sample space • Conditional sample space	All population units Subset of units in population	Tended to disregard when assessing risk Used when thinking about risk applying to me

Table 5. Summary of four students' conceptions and reasoning about risk in the study

The availability heuristic. People naturally make use of their experiences in order to make judgements. The availability heuristic is employed when we recall information from family, peers and the media to analyse risk information (Kahneman, 2011). In this study, the students' use of the availability heuristic was characterised in two ways. Firstly, when faced with several situations from which death could arise, students were asked to rank them in order, from the most to the least risky. The disparity between the students' ranks and those of the NSC and the CDC appear to be due, in part, to the students' reliance on memory which brought to mind examples of moderately common or relatively uncommon causes such as suicide and drowning, and propelled them to the top of their lists. Similarly, common causes of death such as falling or from the effect of smoking, not readily reported in the media, came close to the bottom of their lists. However, when the students were asked to consider the reasons for the different R-lines in Figure 1 (see Section 4.4.1), they were able to draw on their contextual knowledge and provide sensible suggestions. This characterisation of the availability heuristic, when students were told that the risk profiles vary, had a beneficial effect on student reasoning.

Critical questions to ask when judging risk information. The students in the study did not appear to consider the critical questions or to search for background information unless specifically asked how to determine risk. In certain instances, they were aware of notions such as time frame but only if it was provided, such as their reasoning with the visual representation of 15-year breast cancer diagnosis (see Figure 1). In other cases they seemed unaware of the importance of the time frame which meant that they were unable to judge how likely it was to die as the result of a fall or from the effects of smoking. The notion of verbal framing was not investigated thoroughly in the study, although for two students it appeared that framing preference had more to do with numbers than with the wording of the question. With regard to the question "to whom does the risk apply?", students were able to judge risks based on the relevant sample space or conditioning event, but again only if this information was provided. This was evident in their responses to the visual representation of risk in Figure 1, particularly when asked to reflect on the corresponding representation for a different group of women. However, when the sample space was not stated explicitly, for example when ranking the risks in Table 1, the relevant sample space did not appear to be considered. Such findings lend credence to the beliefs of researchers such as Gigerenzer (2002) and Watson (2006) who state that by asking critical questions of risk information and being aware of the ways in which risk can be represented, we may be more informed and not so easily manipulated.

Reasoning numerically with risk. When it came to estimating risk, students appeared to consider both *context* and *consequence*. For example, when describing a risky activity in which they had been involved, their assessment of the size of the risk seemed to be based on the severity of the outcome in conjunction with how likely it would be to experience such an outcome. The student who provided the unicycling example (see Section 4.3) noted that the risk of an adverse event such as falling off was relatively high (70%), but that there were lots of possible things that could go wrong due to the instability of a unicycle and the likelihood of unstable ground. The student who described the hiking scenario gave a much lower estimate (2%), but explained that although the activity in itself could be considered high risk, the preparatory groundwork and experience of the hike leaders would ameliorate this risk substantially. The idea that the severity of the *consequences* may be influencing the students' risk perceptions more than the probability of an undesirable outcome is in keeping with the findings of Sadique, et al. (2013). However, there was also a tendency for students to overestimate the risk of rare but potentially dramatic events such as suicide, which is indicated by the availability heuristic characterisation, and also by the research of Everitt (1999). With regard to calculating risks numerically, the students in the study were more comfortable working with natural frequencies than with probabilistic information such as proportions which is in line with the research (Gigerenzer et al., 2007; Hoffrage et al., 2000; Martignon, 2014; Peters et al., 2007). Another characterisation of reasoning numerically with risk that was evident in the study was the perception of most of the students that risk has a negative connotation. This finding supports Watson's (2006) belief that interpretation and evaluation of the language of risk, when stated in a social context, is crucial. The fact that Gal (2002) includes a language and context component in his probability literacy framework suggests that when teaching risk, the students' perceptions must be considered.

Sample space. The final characterisation of student conception evident in this study is that of sample space. While consideration of sample space underpinned the characterisation of '*Judgement of risks are influenced by identification of relevant sample space*' in Table 4, it is worthy of a mention in its own right. When asked to assess risk by ranking (see Section 4.2) or by calculating having been provided with numerical information in a probabilistic format (see Section 4.4.2), the students tended to be unaware of the sample space of interest. However, when specifically asked to consider risks for a particular group of individuals or units (see Section 4.4.1), or to estimate risks for which they had no relevant numerical information (see Section 4.2.2), they were either able to integrate knowledge of the relevant conditional sample space or to ask for relevant information about particular characteristics to arrive at sensible conclusions. Sample space is well-recognised as a key concept for learning about probability (Batanero & Sanchez, 2005; Jones, Langrall, Thornton, & Mogill, 1997; Savard, 2014). In light of the findings of this small study, it would appear that sample space, as a concept, should also have prominence in the learning of risk.

Curriculum and risk

As evidenced by the findings of this small exploratory study, the four high-achieving students who participated were proficient at calculating risks and relative risks from two-way tables and were able to reason effectively with different risk representations. However, the reasoning strategies employed by these students did not always lead to appropriate judgments or conclusions. The current practice of teaching risk and probability calculations, with little regard to the wider context, results in students being capable of the mathematical manipulations and uncomplicated interpretations of risk statements. However, integrating more context into the situations provided to students, and building in students an awareness of intuitions and potential biases, may result in a deeper conceptual understanding of risk.

Implication for Teaching

Gigerenzer (2002) describes three important steps when it comes to teaching people to reason with risk. The first step, which he calls "Franklin's Law", is to develop the awareness that uncertainty is a given since, according to Benjamin Franklin, "In this world nothing is certain but death and taxes". He recommends demonstrating this through the use of everyday contexts in which uncertainty is present rather than through the use of imaginary situations. The second step, "Beyond ignorance of risks", involves teaching people to know how to use information to estimate risks and to be cognisant of the difficulties that may arise in doing so, such as biased media reporting. The third step, "Communication and Reasoning", includes educating people in the various ways in which risk can be represented, and to develop awareness of the ways in which we can be manipulated by different representations.

With an awareness of the student conceptions that were apparent in this study, we conjecture that attending explicitly to notions such as those outlined in Table 4 may benefit school students such as those who were involved in the study as they learn to interpret and reason with risk in varied situations. Our findings suggest that intuition guided the students' reasoning processes in many of the tasks in this study, particularly when reliable information was not provided. In such instances the students adopted a *risk as feelings* approach. However, when specifically asked for numerical estimates, or for reasons to explain certain factual findings, they took a logical approach, thereby employing *risk as analysis* (c.f. Slovic & Peters, 2006).

References

Anderson, E. E., & Iltis, A. S. (2008). Assessing and improving research participants' understanding of risk: Potential lessons from the literature on physician-patient risk communication. *Journal of empirical reseach on human research ethics: An international journal, 3*(3), 27-37.

Aronson, J. K. (2006). Medicines and the media. *British Journal of Clinical Pharmacology, 61*(2), 121-122.

Batanero, C., & Sanchez, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 241-266). New York, NY: Kluwer/Springer Academic Publishers.

Biggeri, L., & Zuliani, A. (1999). The dissemination of statistical literacy among citizens and public administration directors. *Proceedings of the International Statistical Institute 52nd session.* Helsinki, Finland. Retrieved from http://iase-web.org/documents/papers/isi52/bigg0981.pdf.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in Psychology, 3*(2), 77-101.

Dargahi-Noubary, G. R., & Growney, J. S. (1998). Risk - a motivating theme for an introductory statistics course. *The American Statistician, 52*(1), 44-48.

Dupont, W. D., & Plummer, W. D. (1996). Understanding the relationship between relative and absolute risk. *Cancer, 77*(11), 2193-2199.

Edwards, A. G., Elwyn, G. J., Covey, J., Mathews, E., & Pill, R. (2001). Presenting risk information: a review of the effects of "framing" and other manipulations on patient outcomes. *Journal of Health Communication, 6*, 61-82.

Eppler, M. J., & Aeschimann, M. (2009). A systematics Framework for Risk Visualization in Risk Management and Communication. *Risk Management, 11*(2), 67 - 89.

Everitt, B. S. (1999). *Chance Rules: An Informal Guide to Probability, Risk and Statistics.* New York, NY: Springer-Verlag.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Janlovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy Without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making, 27*, 672 - 680.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review, 70*(1), 1-25.

Gal, I. (2005). Towards "Probability Literacy" for all Citizens: Building Blocks and Instructional Dilemmas. *Exploring probability in school: Challenges for teaching and learning* (pp. 39 - 63). New York, NY: Kluwer/Springer Academic Publishers.

Gardner, D. (2008). *Risk: The Science and Politics of Fear.* Toronto, Canada: McClelland & Stewart Ltd.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you.* New York, NY: Simon & Schuster.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks. *British Medical Journal, 327*, 741-744.

Gigerenzer, G., Gaissmeier, W., Kurz-Milcke, E., Schwartz, L., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological science in the public interest, 8*(2), 53 - 96.

Greer, B., & Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: historical, cultural, social and political contexts. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 297-324). New York, NY: Springer.

Harrabin, R., Coote, A., & Allen, J. (2003). *Health in the News: Risk, Reporting and Media Influence.* London: King's Fund.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating Statistical Information. *Science New Series, 290*(5500), 2261 - 2262.

Iman, R. L. (1994). *A data-based approach to statistics.* Belmont, CA: Duxbury.

Johnson, E. J. (2004). Rediscovering Risk. *Journal of Public Policy and Marketing, 23*(1), 2-6.

Jones, G. A. (Ed.) (2005). *Exploring probability in school: Challenges for teaching and learning.* New York, NY: Kluwer/Springer Academic Publishers.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*, 101-125.

Jones, R. B. (2012). *20% chance of rain: exploring the concept of risk.* Hoboken, NJ: Wiley.

Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY: Allen Lane.

Martignon, L. (2014). Fostering children's probabilistic reasoning and first elements of risk evaluation. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting plural perspectives* (pp. 149-160). Dordrecht, The Netherlands: Springer.

National Safety Council, (2006). *National Safety Council.* Retrieved from Odds of death due to injury: http://www.nsc.org/news_resources/injury_and_death_statistics/Documents/Odds%20of%20Dying.pdf

Peters, E., Hibbard, J., Slovic, P., & Dieckmann, N. (2007). Numeracy Skill and the Communication, Comprehension, and Use of Risk-Benefit Information. *Health Affairs, 26*(3), 741 - 751.

Sadique, M. Z., Devlin, N., Edmunds, W. J., & Parkin, D. (2013). The effect of perceived risks on the demand for vaccination: Results from a discrete choice experiment. *PLOS One, 8*(2).

Savard, A. (2014). Developing probabilistic thinking: What about people's conceptions? In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 283-298). Dordrecht, The Netherlands: Springer.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science, 15*(6), 322-325.

Slovic, P., & Weber, E. (2002). Perception of risk posed by extreme events. *The Conference on Risk Management Strategies in an Uncertain World*, Palisades, NY, 1-21. Retrieved from https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/5461/perception%20of%20risk.pdf

Sorensen, L., Gyrd-Hansen, D., Kristiansen, I. S., Nexoe, J., & Nielsen, J. B. (2008). Laypersons' understanding of relative risk reductions: Randomised cross-sectional study. *BMC Medical Informatics and Decision Making, 8*, 31-37.

Spiegelhalter, D. (n.d.). *2845 ways to spin the Risk.* Retrieved from Understanding Uncertainty: http://understandinguncertainty.org/node/233

Svenson, O., Fischhoff, B., & MacGregor, D. (1985). Perceived driving safety and seatbelt usage. *Accident Analysis and Prevention, 17*(2), 119-133.

Thirlaway, K. J., & Heggs, D. A. (2010). Interpreting risk messages: Women's responses to a health story. *Health, Risk & Society, 7*(2), 107-121.

Thompson, S. C., Armstrong, W., & Thomas, C. (1998). Illusions of control, underestimations, and accuracy: A control heuristic explanation. *Psychological Bulletin, 123*, 143-161.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Utts, J. M. (2005). *Seeing Through Statistics.* Belmont, Canada: Thomson Brooks/Cole.

Watson, J. M. (2006). *Statistical Literacy at School: Growth and Goals.* New York, NY: Routledge.

Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science, 246*(4935), 1232-1233.

Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2010). A demonstration of "less can be more" in risk graphs. *Medical Decision Making, 30*, 661-671.

Judgment of Association between Potential Factors and Associated Risk in 2x2 Tables: A Study with Psychology Students

Carmen Batanero
University of Granada, Spain
Gustavo R. Cañadas
University of Granada, Spain
Carmen Díaz
University of Granada, Spain
Maria M. Gea
University of Granada, Spain

Abstract: This study was aimed to evaluate the accuracy and strategies used in the estimation of association between potential factors and associated risks when data are presented in 2x2 tables. A sample of 414 undergraduate Psychology students from three different Spanish universities was given three different tasks (direct and inverse association and perfect independence) where they had to estimate such association. Most participants judged association in the task where there was perfect independence, but the data contradicted the students' previous expectations. The estimation of association was consistent with the perception of association and the accuracy of estimates increased with correct strategies. Our participants performed worse than secondary school students in a previous study (Batanero et al., 1996) and we found no difference in the three participating universities. We classify the students' strategies in the tasks in levels of complexity and explain the incorrect strategies using the idea of semiotic conflict.

Keywords: risk factors, association, 2-way tables, semiotic conflicts, psychology students.

Introduction

In the past decades we have observed an increasing interest towards risk perception and management in many organizations, such as private companies, banks, hospitals, or schools. Slovic (2000) described a risk as an uncertain event or condition that when happening has a positive or negative effect on a person or a group of people. Gigerenzer (2003) suggested that such an event should be associated with a probability or a frequency with is based on empirical data about its past or potential occurrence. In this definition the author includes different interpretations of probability; such as subjective probability (degree of belief of the person assigning the probability), propensity (physical property of an object; for example, regularity in a dice) or frequentist estimation (information from a large number of observations of the event). Consequently, an event could be perceived or not as a risk, depending on the person's conception of probability.

Today an increasing number of events are described in terms of risk using mathematical formats, such as probabilities, proportions or percentages. The underlying concepts have to be learned in school and for this reason, mathematics educators are becoming interested in students' perception and understanding of risk (Martignon, 2014). In the same way, the assessment of students' potential biases, wrong strategies and misconceptions when interpreting risks is a relevant area of research in mathematics education (Nunes, & Bryant, 2011).

In health or clinical contexts, including psychological evaluation or diagnosis, risk is synonymous of hazards and dangers; for example an illness or the undesirable effect of a treatment (Power, 2007). Risks in these contexts are often associated to decision making in such a way that it is

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 347-363

impossible to make a 'risk-free' decision, unless we leave some potential risk factors unmanaged. A sound understanding of risk and of the associated numerical information is essential for these professionals, in order to make adequate decisions. However, numerous examples that this understanding is not complete are described in the literature related with risk perception and management. For example, Gigerenzer et al. (2007) termed *collective statistical illiteracy* the fact that many professionals have difficulties in interpreting health statistics and draw wrong conclusions in the clinical practice without noticing. Gigerenzer and Edwards (2003) suggest that misperception of risk factors occur by confusing single event probabilities, conditional probabilities, and relative risks. They also suggest that the situation can be improved by education and by representing the information regarding risks in ways that are transparent for the human mind, such as natural frequencies, tree diagrams or two-way tables.

In fact, two-way tables are a main representational tool for bivariate data, and are often used in professional journals to report the influence of risk factors on different pathologies. Its understanding is related to risk perception (Nunes, & Bryant, 2011). Adi, Karplus, Lawson, and Pulos (1978) suggested that providing subjects with information already organised in tables improves their performance in tasks in which they are asked to assess whether there is an association between two events.

These tables; in particular 2x2 tables (two-way tables with only two columns and two rows) are also an important tool in diagnosis and psychological evaluation, where psychologists are confronted with different potential risk factors that may be associated or not with a disorder (risk) (Diaz, & Gallego, 2006). The estimation of association in these tables is the first step to determine which factors are associated to specific risks. However, even when association judgment is a priority learning issue in statistics courses (Zieffler, 2006), little attention is paid to its teaching, in assuming that the interpretation of 2x2 tables is easy. Contrary to this belief, previous research on this area described in Section 2 suggests that people's performance in judging association from data presented in a 2x2 tables is often inaccurate.

This study was aimed to evaluate the accuracy and the strategies used in the estimation of association in 2x2 tables by psychology undergraduate students. A sample of 414 students of Psychology in three different Spanish universities was given three different tasks (where data show direct and inverse association and perfect independence). We compare the judgment of association and the accuracy in the estimation of the strength of association by item and university. The strategies used by the students to judge the existence of association are classified in five levels of complexity described by Pérez-Echeverría (1990). Our results are compared with a previous study with high school students by Batanero, Estepa, Godino and Green (1996). We use the idea of *semiotic conflict* introduced by Font, Godino and D'Amore (2007), to explain the prevalence of incorrect strategies. Some final implications for teaching 2x2 tables are included.

Background

Perception of Risk

The interest in analyzing people's perception of risk arose by the observed fact that people often disagree about the likelihood that a risk happens in presence of some risk factors. Several theories have been proposed to explain why different people make different estimates of the likelihood of risks. Early work in Psychology (e.g., Tversky, & Kahneman, 1974) suggested that people use cognitive *heuristics* in sorting and simplifying information in uncertain situations which lead to biases in decision making. This research also suggested possible explanation of peoples' biases in decision making in terms of heuristics, such as *representativeness, availability or anchoring,* some of which may affect people's perception of risk. For example, the representativeness heuristics occurs when people judge the probability of an event by considering only its similarity with a population. An associated bias is the *insensitivity to prior probabilities,* where people do not take into account a condition that affects the probability for such event. In another example, the *illusion of validity* is the belief that irrelevant information (e.g., personal experience, salience of information) generates additional relevant data for predictions, even when this is not the case.

Association in 2x2 Tables

Between the complete certainty of some events and the complete uncertainty lies the world of association. Association between some variables (factors) and uncertain events (risks) are frequent in many situations both in professional and everyday life. Correct perceptions of these associations help people evaluate the probability of a particular risks taking place in presence of an associated factor.

Correlational reasoning is an important research area which is aimed to describe people's accuracy and strategies to evaluate the presence and strength of association between variables (see Adi, Karplus, Lawson, & Pulos, 1978, for a survey). According to this research, correlational reasoning is an important component of scientific reasoning, as it is a tool for understanding the past, controlling the present and predicting the future (Alloy & Tabachnik, 1984).

Research on correlational reasoning has been often carried out with dichotomous variables; where we use the word *associatio*n to describe the relationship between these variables (the word correlation is used to describe relationships between numerical variables). In this research participants are given some tasks where they are asked to decide if a variable with two modalities (*A,* not *A*) is associated or not with another variable with two modalities (*B,* not *B*).

The pioneer work in this line of research was due to Inhelder and Piaget (1955), who conceived association as the last step in the development of probabilistic reasoning. They performed interviews with 13- 14 year olds students and proposed them the study of the association between eyes colour (blue and dark eyes) and hair colour (fair and brown hair) using data that were formally equivalent to Table 1. According to Nunes and Bryant (2011), understanding the association between the two variables in Table 1 makes three demands on people's reasoning: a) they need to understand that the situation involves randomness, since not always *A* and *B* appear together; b) it is important to recognise the cases (*a* and *d*) that support an association between the variables and the cases (*b* and *c*) that go against the association; and c) the subjects should be able to quantify and compare the positive and negative cases in order to assess whether the co-occurrence observed is not due to chance.

	A	Not *A*	Total
B	a	b	a+b
Not *B*	c	d	c+d
Total	a+c	b+d	a+b+c+d

Table 1. Scheme of a 2x2 table.

Inhelder and Piaget (1955) described different types of strategies used by the children in their interviews to decide if there is association between the variables or not. In a first stage (13 year olds or younger children) the subjects only analyze the favourable positive cases in the association (cell *a* in Table 1). In a second stage of development they compare *a* with *b* or compare *c* with *d.* Inhelder and Piaget suggested that, although these children can compute single probabilities, understanding association requires considering quantities *(a+d)* as favourable to the association and *(b+c)* as opposed to association, and also that it is necessary to consider the relation:

$$R = \frac{(a+d)-(b+c)}{a+b+c+d}$$

where R represents the difference between cases confirming and opposed to the association compared to all the possibilities. We can observe that R=1, when all the data fall in cells a and d; R=-1 when all the data fall in cells b and c and R=0 when $(a+d)= (b+c)$. According to Piaget and Inhelder, recognition of this fact only happens at 15 years of age.

The study by Inhelder and Piaget inspired some further research, and was extensively repeated with subjects at different ages and this research showed that adults have poor correlational reasoning. For example, Smedslund (1963) found adults who based their judgment only on the frequency in cell a in Table 1 or by comparing the frequencies in cells a and b. Allan and Jenkins (1983) showed the tendency to base the association judgments on the difference between the frequency of cell a (cases where A and B happen simultaneously) and frequency in cell d (where neither A or B happen). They remarked that these subjects do not understand that a high frequency in cell d is favourable to a positive association. Jenkins and Ward (1965) pointed out that even the strategy of comparing the probabilities of the table diagonals $(a+d)$ and $(c+b)$ (computing R) considered correct by Piaget and Inhelder is not always valid; if the difference in the totals in the rows or columns in the table is high, this strategy can produce errors. Nevertheless, in Allan and Jenkins's (1983) research this strategy was widely used by adults.

According to Jenkins and Ward the correct general strategy requires comparing conditional probabilities; for example, comparing the probability of B in the group of people with A and in the group of people with not A:

$$D = P(B \mid A) - P(B \mid notA) = (\frac{a}{a+c}) - (\frac{b}{d+d})$$

Pérez-Echevarría (1990) classified all the strategies that have been identified in judging association in 2x2 tables in 5 different levels of complexity: In levels 0 to 3 people only use data in 0 to 3 different cells; in Level 4 the subjects base their judgment of association on additive comparisons of the data in the four cells. In Level 5 the subjects base their judgment of association on multiplicative comparisons between the four cells. These level 5 strategies are the only correct strategies in general; although in some particular tables the subjects can succeed with lower level strategies.

Variables that Affect Perception of Association

Other researchers have focussed on the evaluation of people's accuracy when they estimate the strength of association in 2x2 tables. A typical task in this research is asking people to provide a number between 0 (perfect independence) and 1 (perfect association), according to their perception of their strength of association in the table data. The number provided by the subjects is compared with a statistical index of association in the data (for example, with the Pearson's Phi coefficient).

The main aim of this research is to find the variables that affect the accuracy in the estimation. Crocker (1981) showed that this accuracy is higher when the cell frequencies are smaller; and also when the events co-vary simultaneously through time; for example, when we consider hair and eye colour of the same subjects. Erlick and Mills (1967) found that negative association is estimated as close to zero. Arkes and Harkness (1983) indicates that the frequency in cell a has the greatest impact on the accuracy of estimates. The accuracy of estimation is higher in causal contexts. According to Barbancho (1992), an association between variables may be explained by the existence of a unilateral cause-effect relationship (causal context). However we may find association in the case of interdependence (each variable affects the other), indirect dependence (there is a third variable affecting the other two) or spurious covariation (the association happens by chance). In addition to the estimate accuracy, understanding association involve the discrimination of these types of relationships between variables.

Biases and misconceptions in judging association

Judging association is also subjected to cognitive biases: *Illusory correlation* takes place when people perceive a relationship between variables that do not exists, stronger than existing of opposite to the data association:

> The report of a correlation between two classes of events which, in reality, (a) are not correlated, (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction from that which is reported (Chapman, 1967, pp. 151).

We could consider illusory correlation as a particular instance of the illusion of validity described by Tversky and Kahneman (1974). Many researchers have found that illusory correlation is common and influences the estimates of association (Jennings, Amabile, & Ross, 1982; Wright & Murphy, 1984; Meiser & Hewstone, 2006). The estimates of association are more accurate if people have no prior expectations about the type of association in the data. If the subject's prior expectations agree with the type of association reflected by the empirical data, there is a tendency to overestimate the association coefficient. But when the data do not reflect the results expected by these personal theories, the subjects are often guided by their beliefs, rather than by the data.

Other biases have been found in previous research. Batanero, Estepa, Godino, & Green (1996) analyzed the strategies in association judgments by 213 high school students (16-17 year-olds) and defined different conceptions of association: (a) *causal conception* according to which the subject only considers association between variables, when it can be explained by the presence of a cause - effect relationship; (b) *unidirectional conception*, where students does not accept an inverse association, (c) *local conception,* where the association is deduced from only a part of the data, and determinist conception, when the student considers association only in case of functional (perfect) dependence.

Mathematical Objects Linked to Two-Way Tables and Potential Semiotic Conflicts

Although previous research related to judgment of association in 2x2 tables is abundant, none of this research questioned the students' capacity to interpret the different mathematical objects implicit in the analysis of Table 1, which is a complex semiotic object (Estrada, & Díaz, 2006). Data in cells *a, b, c, d* refer to joint absolute frequencies, each of them for a double condition (values of row and column); however; as suggested by Inhelder and Piaget (1955), their meaning is non-equivalent. A high frequency in cells *a* (presence of character A; presence of character B) and *d* (absence of A, absence of B) would indicate a positive association between the variables; and a high frequency in the other two cells suggests a negative association. Moreover, from a given cell we can deduce joint as well as row and column conditional relative frequencies. For example, from cell *a,* we can compute three different relative frequencies (or three different percents if we multiply each of these relative frequencies by 100):

- Joint relative frequency: $\frac{a}{a+b+c+d}$
- Relative frequency as regards the row total: $\frac{a}{a+b}$
- Relative frequency as regards the column total: $\frac{a}{a+c}$
- Furthermore, we can compute the relative marginal frequencies of rows and columns:

$$\frac{a+b}{a+b+c+d} \text{ and } \frac{a+c}{a+b+c+d}$$

All these mathematics objects co-exist and may be confused by students. In our analysis we use the construct *semiotic conflict* taken from Font, Godino and D'Amore (2007). These authors adapted from Eco (1979) the idea of *semiotic function* or correspondence between an expression and its content and suggest that, in mathematical practices many objects intervene: problems, actions, concepts-definition, language properties and arguments, any of which could be used as either expression or content in a semiotic function. The authors termed *semiotic conflict* any disparity or difference of interpretation between the meanings ascribed to an expression by two subjects.

Consequently, we assume that some incorrect strategies used in judging association maybe explained by misinterpretation or confusion between the different mathematical objects involved in the analysis of 2x2 tables. For example Falk's (1986) described the *fallacy of transposed conditional* by which students do not adequately discriminate the two different conditional probability $P(A|B)$ and $P(B|A)$. Einhorn and Hogarth (1986) observed that some students misinterpret the conjunction "and" when computing probabilities and confuse joint and conditional probability. It is possible that these errors appear when interpreting a 2x2 table and this led the students to use an incorrect strategy to judge association.

Wiht this conjecture in mind, our research was aimed to assess the accuracy in the estimation of association in 2x2 tables by Psychology students. We also try to complement the analyses of strategies made by Batanero et al. (1996) by identifying students' semiotic conflicts when judging association.

Method

The sample included 414 students in their first year of Psychology studies from three Spanish universities: Almeria (115 students), Granada (237 students) and Huelva (62 students), all of them taking an introductory statistics course. The questionnaire was given to the student as a part of a practice. The samples included all the students enrolled in the course and attending the session; the difference in sample sizes was due to the size of the University: Almeria with 2 groups of students, Huelva with 1 group of students and Granada with 4 groups of students. Though the students had not yet studied association in the course they were following; however, they had studied descriptive statistics and probability in the first semester.

The questionnaire was adapted from Batanero, Estepa, Godino and Green (1996) (see Appendix). The context was changed to psychological diagnosis in two items (1 and 2). The frequencies in the table cells were increased in items 2 and 3, since in the original questionnaire the small sizes made the application of the Chi-square statistics invalid. The sign and strength of association were maintained in all the items. The following task variables (Table 2) were considered in the questionnaire:

1. *Sign of association:* We include the three possible cases: direct and inverse association and independence.

2. *Strength of association,* which was measured by the Pearson's Phi coefficient. An item with perfect independence and two items with moderate-high association were included.

3. *Agreement between association in the data and previous theories* suggested by the context. There was one item where the empirical association matched the prior expectations and one where it contradicted the expectations and another with a neutral context suggesting no previous theories.

4. *Type of covariation.* We used three categories of Barbancho's (1992) classification: unilateral causal dependence, interdependence and indirect dependence.

	Item 1	Item 2	Item 3
Sign of association	Independence	Inverse	Direct
Association coefficient (Pearson's Phi)	0	-0.62	0.67
Agreement with prior theories	No	Yes	There is no theory
Type of covariation	Interdependence	Causal unilateral	Indirect dependence

Table 2. Task variables in the items

A qualitative analysis of students' responses served to define three different variables:

1. In part (a) of each item, students are asked to provide an association judgment. We considered three different categories in their responses: (a) the student considers that the variables in the item are related (association); (b) the student considers that the variables are not related (independence); and (c) the student does not decide (no judgment).

2. The estimation of the strength of association is deduced measuring the exact position of the point drawn by the student on the numerical scale (0-1) in the second part of the item.

3. Finally, a qualitative analysis of the reasoning given by the students to justify their responses serves to identify the strategies used by the students and their semiotic conflicts. The classification of strategies was performed independently by two different members of the team; in the case of disagreement, it was revised by other team members until an agreement was reached.

Perception of Association

Association Judgment

In order to analyse the students' competence to perceive the possible association between the data given in each item, we present the percentage of students who accepted the existence of a relationship between the variables in Table 3. In the last columns we add the value of the Phi association coefficient for the data and the relationships between prior theories and data.

	Almería (n=115)	Granada (n=237)	Huelva (n=62)	Total (n=414)	Association coefficient	Prior theories vs data
Item 1	87 (75.7)	194 (81.9)	42 (67.7)	323 (78.1)	0	Do not agree
Item 2	108 (93.9)	230 (97.1)	60 (96.8)	398 (96.1)	-0.62	Agree
Item 3	107 (93.1)	226 (95.4)	53 (85.5)	386 (93.2)	0.67	No theories

Table 3: Frequency (and percent) of students considering there is association in the data.

Most students indicated the existence of association in all items, in particular when the association was confirmed by the data, but also in Item 1 (where the data correspond to perfect independence). We explain this result by the mechanism of illusory correlation (Chapman, 1967), since in this item data contradicts the students' previous expectation (that stress is related to insomnia). Many students expressed their previous belief that both variables should be related in this item in their answers: "*There should be some relationship, since in my experience stress due to family or other type of problems may be a cause of insomnia*" (Student 213). *"In my opinion insomnia and stress are related, since most people who have insomnia suffer from stress"* (Student 5), or "*Yes, because people with insomnia do not rest well and this causes extra stress that is added to stress due to other external factors*" (Student 42).

These responses and the high percentage of students that argued that the variables are associated (in spite of perfect independence in the data) suggest that these previous theories affect the student perception of association. Moreover these responses show that part of these students might identify correlation and causation. Our results in this item are worse than those in Batanero et al. (1996) with secondary school students, since these authors only found 55.4% of students judging association in an item with exactly the same data and a similar context (smoking and bronchial disease).

Our students outperformed in Item 2 (inverse association) those in Batanero et al. (2006), where only 47.1% of students considered association. This result could be explained by the change in context (diet and digestive troubles in Batanero et al.) and the increased sample size in our item. Consequently, the unidirectional conceptions of association described by Batanero et al. (1996) by which students perceive negative association as close to independence hardly appear in our research. Results in Item 3 were very close in both studies (92.5% in Batanero et al., 1996), where we only increased the frequencies without changing the context or the strength of association.

In Table 3 we observe little differences in the percentages of students accepting association in each item in the three universities. In order to check the homogeneity of these samples we carried out a Chi-square test of homogeneity between the three samples and we obtained the following results (Chi= 0.99; 6 d.f.; p=0.9861). This result led us consider the subsamples to be homogeneous as regards the percentage of students considering association in each item and that there was no much difference in this variable, in spite of the different educational context.

Estimating the strength of association

In the second part of their response to each item, the students provided a score between 0 and 1 according to the association strength they perceived in the data. This score is an estimate of the association coefficient (disregarding the sign). In Table 4 we present the mean score obtained by university and in the whole sample by item, as well as the characteristics of each item. The most accurate estimate corresponds to Item 3, where the students showed no prior expectations: The average estimate is very close to the empirical coefficient in all the samples and in the whole sample.

	Mean estimate				Association coefficient	Prior theories vs data
Item	Almeria (n=115)	Granada (n=237)	Huelva (n=62)	Total (n=414)		
1	0.51	0.47	0.44	0.47	0	Do not agree
2	0.78	0.72	0.73	0.73	-0.62	Agree
3	0.75	0.68	0.68	0.70	0.67	No theories

Table 4. Mean estimates of the association coefficient

In Item 1, which corresponds to perfect independence, the global mean value was 0.47, and about the same value in each sample. As we described in the previous section, the students' previous expectations contradicted the independence in the data in this item. Moreover, in this item cell *a,* (simultaneous presence of both stress and insomnia) contains the maximum absolute frequency; according to Arkes and Harkness (1983) cell *a* has the greatest impact on the subject's attention,.

The estimate for Item 2 (inverse dependence), was higher than the empirical value association in all universities. Thus, in our students we did not find a significant presence of the unidirectional conception described by Batanero et al. (1996). In this item the students' previous expectations coincided with the association in the data. Some students expressed these beliefs in their response: "*I think there is relationship between having brothers or sisters and being a problematic child, since children having brothers or sisters are raised with different moral values, such as generosity or empathy, than only children*" (Student 44); "I*n my experience only children are troublesome"* (Student 172).

Moreover, both in the whole sample and in each university most students indicated that there was association in this item (the sign of association was not requested).

This judgment of association was consistent with the students' estimation of the association coefficient (Table 5) with the differences in the mean estimates in students judging association in the item or not judging association being statistically significant in the t-test of differences for all the items.

	Association	No association				
Item	Mean estimate	Mean estimate	Mean difference	Standard error for the mean difference	Students t	p-value
1	0.56	0.18	0.38	0.02	17.00	0.000
2	0.75	0.31	0.44	0.05	8.76	0.000
3	0.73	0.22	0.51	0.04	12.10	0.000

Table 5. Mean estimates of the association coefficient in students considering or not association in the data and t- test of differences in means

The students from Granada and Huelva estimated an average lower association in all the items than the students from Almeria; however, the differences in mean estimates were not statistically significant in the one-way ANOVA tests (F=0.487, p=0,615 for Item 1; F=0.260, p=0,771 for Item 2 (F=0.308, p=0.735 for Item 3 with d.f=2,411). These results suggest that we cannot reject the hypothesis that the students' average estimates are similar in the three universities, despite the possible differences in educational context. Consequently we decided to combine the three samples for further analyses.

Students' Strategies

Level of strategies

To complement the above results we performed a qualitative analysis of students' strategies. Firstly, we differentiated three groups of strategies:

Correct strategies. These always produce a correct association judgement, such as comparing conditional frequencies by rows or columns. An example is comparing the proportion of people with stress disorder with those with no stress disorder in people with and without insomnia in Item 1.

Partly correct strategies. Procedures that provide a correct association judgment for the particular table in which it is applied, but that is not valid in general for any 2x2 table. An example in Item 2 is comparing the frequencies in the first row; as we found less problematic children among those who have sisters or brothers, the student correctly deduces association in the data. The procedure works for this example, but not for the general case; for example, however, this procedure leads to an incorrect judgment in Item 1.

Incorrect strategies. When students use a procedure that is incorrect in all type of tables. A particular example is not using the data table; another example is using only one cell, as in this case it is impossible to study the variation of frequencies with different combinations of the variables.

This classification was crossed with the levels of difficulty proposed by Pérez-Echeverría (1990). All level 0 and 1 strategies are incorrect, while part of the strategies in upper levels are partly correct or correct.

Level 0 Strategy. The student uses no data from the table and only takes into account his/her own previous beliefs about the association that should be in the variables; the illusory correlation (Chapman & Chapman, 1969) is visible in these students; for example: "*The variables are related, since when you do not sleep your stress increases*" (Student 5).

Level 1 Strategy. When the student only used one cell in the table. Usually the student use only cell *a* because in this cell when both characters are present and so the frequencies in this cell have a higher impact on our attention (Smedlund, 1963; Beyth -Marom, 1982; Shaklee & Mins, 1982, Yates & Curley, 1986): *"there is association, since 90 people (36 % of the sample have insomnia and stress"* (Strategy 1.1; Student 111). Other students used one of the cells *b* or *c* that contradicts the association: *"There is no relation since there are 60 people with stress and no* *insomnia and this is a big percentage"* (Strategy 1.2, Student 51). These students expect a deterministic relation between the variables as they assume that there could be no exception in the association; that is, he expected that all people with stress should suffer from insomnia.

Level 2 Strategy. Some students use two cells; for example, they compare *a* with *b* or *a* with *c*. Consequently they deduce the association from only one conditional distribution in the table: *"If you look to the people with insomnia, there are more people with stress (90) than without stress (60)"* (Strategy 2.1, Student 21). In this item this strategy is incorrect, as it led to an incorrect judgment; however, if used in items 2 or 3 is partly correct as it will produce a correct judgment. Other students compared the cells with maximum and minimum frequency: *"There are 90 people with stress and insomnia and 40 without stress and without insomnia; 90>40, but the relation is not too strong"* (Strategy 2.2, Student 61). This strategy is incorrect, because cell *d* is also relevant for the association, contrary to what the student assumes.

Level 3 Strategy. In this strategy the student uses three cells; for example, he compares cell *a* with *b* and *c*. In general, these students discarded cell *d* that corresponds to the absence of both characters: *"There is relationship as there are more people with stress and insomnia (90 people) and exactly the same number (60) with either stress and no insomnia or insomnia and no stress"* (Strategy 3.1, Student 153). The strategy could work with items 2 and 3 and then is partly correct for these items but it is incorrect for Item 1.

Level 4 strategies. These strategies are based on additive comparisons of the four cells. One example is comparing the sum of diagonals *(a+d)* with *(b+c)*: *"There are 130 people with both stress and insomnia or no stress and no insomnia, while there are only 120 with one of these symptoms.* (Strategy 4.1, Student 176). This strategy was found by Allan and Jenkins (1983) and could provide a good solution when the marginal frequencies (number of people with and without insomnia) were equal, according to Shaklee (1983. As we see in the example, it does not work with Item 1; for this reason we consider this strategy to be incorrect for this item. In another example, students compared two conditional distributions in an additive way: *In people with insomnia there is a difference of 30 having stress, while the difference in people without insomnia is smaller (20)"* (Strategy 4.2, Student 267). Finally other students compared all the absolute frequencies among them: "*There are many people with stress and insomnia (90) but the relationships is not strong, since having stress and no insomnia or insomnia and no stress (60) is also high, much higher than no insomnia and no stress (40)* (Strategy 4.3, Student 156).

Level 5 strategies. Some students use all the four cells with multiplicative comparisons, but still may be incorrect or partly correct. For example, a wrong strategy is to compute all the joint relative frequencies and compared them: *"I computed the percent of each data and compared the results:* $\frac{90}{250}100 = 36\%$; $\frac{60}{250}100 = 24\%$; $\frac{60}{250}100 = 24\%$; $\frac{40}{250}100 = 16\%$ "(Strategy 5.1, Student 11). This procedure is incorrect, because the association should be deduced from conditional distributions and not from joint distributions. An example of partly correct strategy is assuming that all joint relative frequencies in the table should be identical (that is, 25% cases in each cell). We considered this strategy partly correct because the student computed some "expected" frequencies, compared them with the observed frequencies and correctly deduced that there was association because these two types of frequencies were different. However the strategy is not valid in the general case.

> I divided 250 between 4 (25%) to see the number of children we should expect in each cell, in case of no relationship. However, although the number of only child who are problematic are close to 25% there is a big difference in the other cells; therefore there should be a relationships (Strategy 5.2, Student 1).

Finally, among the level 5 correct strategies we found students who compared two conditional distributions by row; for example, *a/(a+b)* with *c/(c+d)* or else compared conditional distributions by columns: “*When we observe the table, 60% of people with insomnia have stress and also 60% of people with no insomnia have stress; the percentage is the same*” (Strategy 5.3, Student 28). Another correct strategy is comparing odds in favour and against *B* for each value of *A*; which was described by Batanero et al. (1996): “*There are 90 people with insomnia for every 60 with no insomnia when you have stress; that is the odds are 3/2; the same odds 60/40 apply when you do not have stress*” (Strategy 5.4, Student 21).

In Table 6 we present the frequency of responses in the above categorization and percent of students in the sample. As regards levels of complexity, students tended to use either level 2 or level 4 strategies none of which are correct, although part of them are partly correct and helped the students to get a correct association judgment. There was a big percentage of students who did not use all the cells information, since their strategies were level 3 or lower.

About 26% of students who used level 4 strategies compared joint frequencies among them, an incorrect strategy described by Batanero et al. (1996) and about 74% of them used the four cells with additive comparisons, a partly correct strategy described by Inhelder and Piaget (1955) in the concrete-operation level but that also was found by Batanero et al. (1996) in high school students. Finally most of level 5 strategies were correct as students either compared the odds ratios (relative risks) or compared conditional distributions a strategy proposed by Jenkins and Ward (1965) and also found in Batanero et al. (1996).

Level	Correctness	Item 1 (Independence)		Item 2 (Inverse)		Item 3 (Direct)	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
Level 0	Incorrect	13	3.1	15	3.6	15	3.9
Level 1	Incorrect	73	17.6	20	4.8	33	8.0
Level 2	Incorrect	108	26.1				
Level 2	P. correct			154	37.2	153	37.0
Level 3	Incorrect	27	6.5				
Level 3	P. correct			16	3.9	9	2.2
Level 4	Incorrect	27	6.5	12	2.9	19	4.6
	P. correct	76	18.4	115	27.8	100	24.2
Level 5	Incorrect	20	4.8	29	7.0	25	6.0
	P.correct	10	2.4	8	1.9	6	1.4
	Correct	46	11.1	37	8.9	28	6.8
Do not explain their strategy		14	3.4	8	1.9	25	6.0
Total		414		414		414	

Table 6. Frequency of strategies (and percent of students) by item, level and correction

To conclude, there was a scarce use of correct strategies (11%, 8.9 and 6.8 depending on the item). In spite of this, most students correctly perceived the association in items 2 and 3 and their estimation of association was reasonably accurate. The explanation is that an important percentage of students used partly correct level 2, 3 and 4 strategies in these two items, where these partly correct strategies are used to obtain a correct association judgement. Specifically, 79.7% of students used partly correct strategies in Item 2 and 71.6 % in Item 3, while in Item 3 the percentage was only 31.9%. These different percentages, added to the illusory correlation phenomena explaining the different performance in the three items.

The mean estimate of the association coefficient in according to the strategy correctness (incorrect, partly correct or correct strategy) is presented in Table 6 with the typical error. In the last two columns of this table we present the association coefficient (Phi value) in each item. We observe that the estimated value of the association coefficient is closer to zero when the strategy is more correct; in particular the estimated average value is very close to zero for correct strategies. This result suggest that partly correct and correct strategies help students perceive independence in the data in this item, even when their previous beliefs about the association were contrary to independence.

In the items 2 and 3, the average estimated value of the association coefficient increase with the strategy correctness and this fact suggest that partly correct and correct strategies helped the students perceive that these data presented high-moderate association. The differences in the estimated average values according the type of strategy (incorrect, partly correct or correct) were statistically significant in the Anova test of in all the three items (p values included in the last column in Table 7.

Item	Incorrect strategy		Partly correct strategy	Correct strategy			Association coefficient	p value (Anova)
	Mean	Typical error	Mean	Typical error	Mean	Typical error		
1	0.536	0.012	0.432	0.024	0.174	0.038	0	0.000
2	0.655	0.029	0.743	0.011	0.809	0.025	-0.62	0.001
3	0.665	0.026	0.716	0.013	0.756	0.029	0.67	0.035

Table 7. Mean estimate of association coefficient in different type of strategy

Semiotic Conflicts in Interpreting 2x2 Tables

The high percentage of students declaring that the variables are associated in Item 1 (78.1%; Table 3), consistently with a moderate-sized average estimated coefficient (.47; Table 4), shows that part of these students reasoned according to illusory correlation. Their incorrect perception of association may also be explained by the wide use of incorrect strategies (64.7) in this item.

To follow we use the idea of semiotic conflict to explain the high percentage of incorrect strategies in Item 1. Our suggestion is that part of the students using these incorrect strategies confused or misinterpreted some of the mathematical objects implicit in 2x2 tables; or else attributed to them some incorrect properties.

To check this conjecture we performed a further detailed qualitative analysis of the students' responses linked to incorrect strategies in Item 1. This analysis helped us to identify these latent semiotic conflicts. To follow we classify these conflicts, according to whether they involve incorrect properties assigned to association or incorrect properties assigned to independence. We include a typical response in each category

Incorrect properties assigned to association. In order to justify association in Item 1, some students assigned non-existent properties to association. The more frequent incorrect properties of association assumed by the students were the following:

•*Identifying association and causality:* Although causality always involves association, association does not always involve causality; but some students misinterpreted that this relation was symmetrical. This belief was also found in Batanero et al. (1996) who assumed it was a stable conception (causal conception of association). In our study this belief appeared in all the level 0 strategies, not only in Item 1, but also in items 2 and 3.

•*Assuming that association can be deduced from only a part of the data.* Association is a property of the data distribution and not a property of some isolated values of the distribution. However, in level 3 or lower level strategies, the students' discard some data in the table to judge the association. Using only a part of the data to solve 2x2 table association problems was also described by Smedlund (1963), Beyth -Marom (1982); Shaklee and Mins (1982), Yates and Curley (1986) and Pérez-Echeverría (1990). Depending on the table particular data Level 2 and 3 strategies may work for the particular problem, as happened in our study. For items 2 and 3 these strategies were partly correct and provided a successful response. However, in Item 1, students who analyzed only a part of the data usually led to a judgment of positive association, as well as to a moderate or high association coefficient.

•*Assuming that there is association when the absolute frequency in the distribution of a variable, when conditioned by the other variable changes.* We can see this belief in those students who deduced association from additive comparisons (Level 4 strategies). This strategy should have been overcome by our students, since according to Inhelder and Piaget (1955) is prior to the formal operations stage. However, it appeared in our sample in Strategy 4.2. Students using this procedure only took into account the favourable cases (and not all the possible cases) when comparing probabilities. Therefore this strategy involves *a conflict* in understanding the idea of probability or confusion between favourable cases and probability. Again the strategy worked well in items 2 and 3, but not in Item 1.

•*Assuming that a difference between the sums of diagonals in the 2x2 table involves association in the data.* This strategy was considered to be correct by Piaget and Inhelder, but Allan & Jenkins (1983) and Shaklee and Mins (1982) suggested it does not work in the general case as we see in Item 1, where Strategy 4.1 lead to an incorrect judgment. However it worked well for items 2 and 3 in our study.

•*Assuming that a>d in the case of association*. This incorrect property appears in Strategy 2.2, where students did not understand that cell *d,* has the same relevance as cell *a* on the association. We did not find this strategy in previous research.

Incorrect properties attributed to independence. In addition to attributing nonexistent properties to association, other students required incorrect properties to decide that there was perfect independence in the data. In particular we observe some students who expected identical joint frequencies in the case of independence (Strategy 5.2). This belief also involves some confusion between the ideas of independence and equiprobability and was not described in previous research.

Discussion and Implication for Teaching

Most psychology students in our study judged association, even in cases where there was none and our results were worse than those in Batanero et al. (1996) in perceiving independence. However our students had a better perception of negative association. These authors did not inform about the estimate of association by their students so that we provide new information about this point. In our study the estimation was very accurate in the case of association and poorer in the case of independence. The estimation was consistent with the association judgment. Results were very close in all participating universities, which suggest that students' reasoning and beliefs as regards association were similar in these different educational contexts.

The illusory correlation phenomenon and the students' previous beliefs may have influenced the students' association judgment and their accuracy in estimating the association coefficient in Item 1 (independence). However, the high percentage of incorrect strategies in this item suggested additional problems in the understanding of the ideas of association and independence. The students tried to justify their previous belief that there was association in this item, using a variety of incorrect strategies. The reasoning behind their explanations revealed a number of incorrect properties assigned to either association or independence that, on one hand served to justify their incorrect strategies and on the other hand confirmed their previous expectation in the association of the data.

We observed the causal conceptions described by Batanero, Estepa, Godino and Green (1996), but not the unidirectional conception, since most students in our study perceived the association when this was negative. In addition we listed some new semiotic conflicts related to misinterpreting mathematical objects and attributing incorrect properties to the ideas of association and independence.

For example, some students assumed that association may be judged from only part of the data, and did not perceive association as a property of the distribution. A high percentage of students tried to deduce the existence of association basing their conclusions only on absolute frequencies. This behaviour was also described by Konold, Pollatsek, Well y Gagnon (1997), who suggested that, when given two groups to compare students rarely used a statistically appropriate method of comparison. In particular, in our study many students tried to compare two groups in the 2x2 table using frequencies rather than percents to make the groups comparison.

Other students rejected association in the data when the frequencies in cells *b* or *c* was not null. They expected a determinist relationship between the variables; however an association between two variables indicates a tendency but we cannot be certain how the association will affect each individual. Even a moderate association allows doctors and psychologists to provide advice to people in danger of some risk; but unless the association is perfect they cannot make exact predictions about what will happen to individuals as a result of a given behaviour (Nunes, & Bryant, 2011).

Finally, we noticed that some students required that the four cells in the table should have equal frequency in case of independence; these students showed an intuitive idea of what is an expected frequency, but however, they confused independence and equiprobability.

According to Schield (2006), an educated person should be able to critically read tables in the press, Internet, media, and professional work. This involve not only the literal reading, but being able to identify trends and variability in the data, including the correct judgment of association. All these reasons and our results suggest the need for further research about teaching association, since the causal conception and the effect of illusory correlation does not seem to improve with traditional instruction (Batanero, Godino, & Estepa, 1998). Since semiotic conflicts do not assume a strong conviction on the part of the students, it is possible to change them with adequate instruction and then the identification of these conflicts in the students is a first step in order to correct their wrong reasoning and improve their competence in judging association.

According to Inhelder and Piaget (1958) and Adi, Karplus, Lawson, and Pulos (1978), understanding association depends on understanding probabilities and proportions. These two types of competencies should be developed in our students so that they can progress in correlational reasoning. Moreover, in order to draw inferences from the frequencies in 2x2 tables, students must understand the relevance of the different cells in Table 1 for a mutual relationship between the variables (Nunes, & Bryant, 2011); although students easily recognise the relevance of cell *a* for the association it is harder that they view the relevance of cell *d*. Nunes and Bryant (2011) suggest that understanding the relevance of call *d* for the association requires reasoning about a contradiction; this is not simple in deterministic situations and less in correlational situations. Moreover these authors also suggest that correlation reasoning requires understanding randomness and assessing the extent to which the association we perceive between the table frequencies departs from what could be expected by chance.

This formation is particularly needed in the statistics education of professionals like psychologist who not only should assess association in risk situations, but only communicate this information to their clients. For example Gigerenzer and Edwards (2003) indicate that even simple probabilities are a steady source of miscommunication because we often leave open the sample space or the population to which the probability refers. The same ambiguity occurs in communicating clinical risk, and the associate factors, such as the side effects of a drug. For this reason 2x2 tables, when properly understood and interpreted may turn in an useful tool for psychologists.

All these abilities should be developed in the students with a careful planning of teaching, Our purpose is to continue this work by designing teaching activities that confront students with their biases and help them overcome them as well as improve their partly correct strategies.

Acknowledgements

Research supported by the project EDU2013-41141-P (MEC, Spain) and group FQM126 (Junta de Andalucía).

References

Adi, H., Karplus, R., Lawson, A., & Pulos, S. (1978). Intellectual development beyond elementary school: correlational reasoning. *School Science and Mathematics, 80*(8), 675-683.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation, 14* (4)*,* 381-405.

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information, *Psychological Review, 91* (1), 112-149.

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112* (1), 117-135.

Barbancho, A. G. (1992). *Estadística elemental moderna* (Modern elementary statistics). 15th edition. Barcelona: Ariel.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. (1996) Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*(2), 151-169.

Batanero, C., Godino, J. D., & Estepa, A. (1998). Building the meaning of statistical association trough data analysis activities. In A. Olivier y K, Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (v.1, pp. 221-242). Stellembosch, South Africa: International Group for the Psychology of Mathematics Education.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and cognition*, *10* (6), 511-519.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6* (1), 151-155

Chapman, L. J. & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs, *Journal of Abnormal Psychology, 74* (3), 271-280.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90* (2), 272-292.

Díaz, J., & Gallego, B. (2006). Algunas medidas de utilidad en el diagnóstico (Some useful measures in diagnosis). *Revista Cubana de Medicina General Integrada, 22*(1). Retreived from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252006000100008.

Eco (1979). *Tratado de semiótica general* (General semiotics). Barcelona: Lumen.

Einhorn, H. J., & Hogart, R. M. (1986). Judging probable cause. *Psychological Bulletin*. *99* (1)*,* 3-19.

Erlick, D. E., & Mills, R.G. (1967). Perceptual quantification of conditional dependency, *Journal of Experimental Psychology, 73* (1)*,* 9-14.

Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores* (Preconceptions on association and its evolution with computer-based teaching). Unpublished Ph.D. University of Granada, Spain.

Estrada, A., & Díaz, C. (2006). Computing probabilities from two way tables. An exploratory study with future teachers. In A. Rossman, & B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics.* Salvador (Bahia): International Association for Statistical Education. Retrieved from: http://iase-web.org/Conference_Proceedings.php?p=ICOTS_7_2006.

Falk, R. (1986). Conditional probabilities: insights and difficulties. In R. Davidson, & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. (pp. 292-297). Victoria, Canada: International Statistical Institute. Retrieved from: http://iase-web.org/Conference_Proceedings.php?p=ICOTS_2_1986.

Font, J. D., Godino, J. D., & D'Amore, B. (2007). An ontosemiotic approach to representations in mathematics education. *For the Learning of Mathematics, 27* (2), 3-9.

Gigerenzer, G. (2003). *Reckoning with risk: learning to live with uncertainty.* London: Penguin.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal, 327*(7417), 741-744.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological science in the public interest, 8*(2), 53-96.

Inhelder, B., & Piaget, J. (1955). *De la logique de l´enfant à la logique de l´adolescent.* (From the child's logic to the adolescent's logic). Paris: Presses Universitaires de France.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of the contingency between responses and outcomes, *Psychological Monographs, 79* (1)*,* 1-17.

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). New York: Cambridge University Press.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997) Students analyzing data: Research of critical barriers. In J. Garfield, & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 151-16). Voorburg, The Netherlands: International Statistical Institute.

Martignon, L. (2014). Fostering children's probabilistic reasoning and first elements of risk evaluation In E. J. Chernoff, B., & Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 149-160). Dordrecht: Springer.

Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology, 363*(3), 315-336.

Nunes, T., & Bryant, P. (2011). Understanding risk and uncertainty: the importance of correlations. *EM TEIA, 2*(2). Retrieved from: www.gente.eti.br/revistas/index.php/emteia/.

Pérez-Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico* (Psychology of probabilistic reasoning). Madrid: Ediciones de la Universidad Autónoma.

Power, M. (2007). *Organized uncertainty: Designing a world of risk management.* Oxford; Oxford University Press.

Schield, M. (2006). Statistical literay survey analysis: reading graphs and tables of rates percentages. In A. Rossman, & B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics.* Salvador (Bahia): International Association for Statistical Education. Retrieved from: http://iase-web.org/Conference_Proceedings.php?p=ICOTS_7_2006.

Shaklee, H. (1983). Human covariation judgment: accuracy and strategy. *Learning and Motivation, 14* (4), 433-448.

Shaklee, H., & Mins, M. (1982). Sources of error in judging event covariations: Effects of memory demands, *Journal of Experimental Psychology Learning, Memory and Cognition, 8*(3), 208-224.

Slovic, P. E. (2000). *The perception of risk.* London: Earthscan Publications.

Smedlund, J. (1963). The concept of correlation in adults. *Scadinavian Journal of Psychology, 4* (1), 165-174.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.

Wright, J. C, & Murphy, G. L. (1984). The utility of theories in intuitive statistics: the robustness of theory-based judgments, *Journal of Experimental Psychology General, 113*(2), 301-322.

Yates, J. F. & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement, *Acta Psicológica, 62* (3), 293-302.

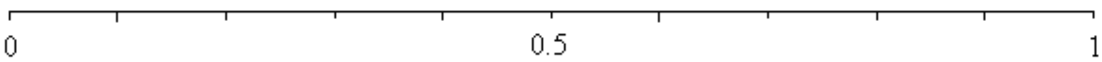
Zieffler, A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course.* Unpublished PhD. University of Minnesota.

Appendix: Questionnaire

Item 1. A researcher is studying the relationship between stress and insomnia. In a sample of 250 people he observed the following results:

	Stress disorders	No stress disorders
Insomnia	90	60
No insomnia	60	40

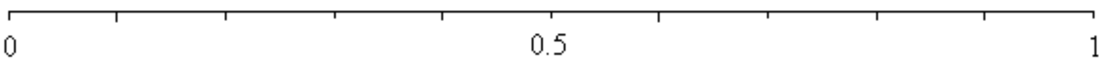
a. Looking to these data, do you think there is a relationship between stress and insomnia? Explain your response.
b. Please mark on the scale below a point between 0 (minimum strength) and 1 (maximum strength), according the strength of relationship you perceive in these data.



Item 2. A psychologist got the following data to study the possible association between being an only child and being a problematic child:

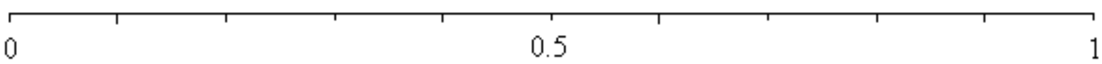
	Problematic child	No problematic child
Having some brothers /sisters	40	100
No brothers or sisters (only child)	100	10

a. Looking to these data, do you think there is a relationship between being an only child and being a problematic child? Explain your response
b. Please mark on the scale below a point between 0 (minimum strength) and 1 (maximum strength), according the strength of relationship you perceive in these data.



Item 3. In order to assess possible association between sedentary life (not performing physical exercise) and allergy the following data were obtained:

	Suffering allergy	Do not suffer allergy
Sedentary life	130	30
No (Active life)	20	120



a. Looking to these data, do you think there is a relationship between sedentary life and suffering allergy? Explain your response.
b. Please mark on the scale below a point between 0 (minimum strength) and 1 (maximum strength), according the strength of relationship you perceive in these data.

Adults' Perceptions of Risk in the Big Data Era

Theodosia Prodromou
University of New England, Australia

Abstract: The present digital era has seen rapid growth in the availability of big data; we were curious about whether such availability of data changes perceptions and assessments of risk. In this paper, we investigate adults' (35-63 years old) perceptions of risk in the big-data era and how it figures in their everyday life. We developed decision-making scenarios for socio-economic, environmental and health topics that involve modelling with personal value systems alongside Gapminder word map data. Going beyond the idea of risk in statistical theory, we attempt to gain an understanding of the processes by which adults assess risks.

Keywords: risk, assessments of risk, big data, informal risk-based reasoning, gapminder.

Introduction

The concept of risk is all-pervasive presence in our daily lives since it permeates decision-making processes in individuals' daily lives, whether personal, professional, or social. Indeed, risk, and the way it is managed, is a critical aspect of decision making at all levels. For instance, we evaluate profit opportunities in business with respect to the engendered countervailing risks, and we make decisions for ourselves or for our family members by analysing the risks involved in areas such as healthcare, sports and exercise, etc.

It is highly significant to make a "risk analysis" when planning any new project or business project and evaluate solutions of types of problems on a risk-cost or cost-benefit basis, such problems might be global, political, financial, and individual. Risk appears to be an extraordinarily diverse area, being studied in a wide range of disciplines. More specifically, the notion of risk is commonly used nowadays in science, medicine, and technology, having its roots in the area of decision-making (Edwards & Tversky, 1967). "Making important decisions in the face of uncertainty is unsettling and difficult and so is a vital area" (Spiegelhalter, 2012). Making decisions about problems from all the aforementioned range of disciplines, demands basing judgements on data to balance effect and likelihood, along with other values-based considerations, of likely hazards.

The word "risk" is hard to pin down (BSON, n.d.); there is still no broad consensus on the meaning of this term. In references to risk, the necessary meaning of the idea of risk is subject of a variety of epistemological perspectives (Adams, 1995; Stirling, 1999; Weale, 2002). Indeed, various national and international standards and guidelines mention risk, and there are many different definitions of the underlying concept. "Even among risk practitioners in the various professional bodies there is an ongoing debate about the subject matter at the heart of their discipline. And of course there is huge variation in the general literature, reflecting the lack of official agreement on the basic definition of risk" (Hilson & Murray-Webster, 2004, p. 2). Nevertheless, all definitions agree that risk is related to uncertainty, and it has consequences (Hilson et al., 2004). Hilson et al., (2004) defines risk in terms of "an uncertainty that could have a positive or negative effect on one or more objectives".

In this paper, I use elements of Hilson's et al. (2004) definitions of risk. I am particularly interested in defining risk as a consequence of uncertainty (Hilson et al., 2004) and relate risk to ideas of probability and likelihood (Pratt, Ainley, Kent, Yogui, & Kapadia, 2011).

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 364-377

In the following section we will discuss the concept of risk in terms of uncertainty and differentiate between risk and uncertainty, building the definition of risk as a consequence of uncertainty during decision-making.

Risk as a Consequence of Uncertainty

We all use personal intuitions to evaluate and interpret “risk” when faced with a new and uncertain event. While there is concern because risk does not appear to underlie people’s intuitive and every day rationalizations (Lupton, 1999), these concerns strengthen when specific events such as property damage or loss after a fire event are brought to the public attention.

The notion of “risk” is inextricably linked to the notion of uncertainty. Risk depends on uncertainty in a potential fact, event, outcome, or scenario, etc. For example, uncertainty causes mortgage issuers to demand property purchase insurance. The person or corporation occupying the mortgage-funded property must purchase insurance on real estate if we intend to lend them money. It would be a risk to lend them money because we did not know if apprehension or dread was about to occur. The future is always unknown, and we are faced with numerous uncertain outcomes (from the mundane—e.g., “we don’t know when it will start raining”—to the extreme—e.g., “we don’t know when or where the hurricane/typhoon will make landfall”).

People’s perceptions of risk arise from perceptions of the quantification of uncertainty. Nevertheless, it is important to differentiate between mathematical uncertainty (i.e., a quantification of probability) and epistemological uncertainty (lack of knowledge about outcomes).

Literature concerned with the relation of risk and uncertainty to ideas of probability and likelihood has attempted various methods to quantify risk and uncertainty. Indeed, both risk and uncertainty are defined by probabilities or based on discrete and continuous probability distributions. Risk isn’t the same as the underlying prerequisite of uncertainty.

My particular concern is about the relation of risk to ideas of probability and likelihood, and statistical interpretation of everyday situations that involve risk. Risk may be quantified using diverse ways, or we may refer to risk without any quantification. For example, the risk of slipping and falling when it is raining is higher than that of slipping and falling when it is a sunny day.

In both official government literature and in media reports, risks reported are often expressed as percentages, or as 1 in *N* cases. So, for example, the risk of cancer disease resulting from drinking, or the risk of cybercrime, or the risk of influenza to air travel is mentioned in terms of mathematical probabilities.

However occasionally increased risks reported by media can be presented more dramatically through the quotation of relative risks. Indeed, the media may state when reporting risk, of, say, exposure to cigarette smoke significantly increases the (relative) risk of developing of lung cancer by 30% in smokers. What exactly does that mean? Say the absolute risk of developing a disease is 10 in 100 in non-smokers. The 30% relates to the 10—so the absolute increase in the risk of 30% of 10 is 3 and the absolute risk of smokers developing lung cancer is 13 in 100.

The media are crucial players in the communication about risk and media often misrepresent risk statistics and report risk involved in some medical cases in a misleading manner (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007). While some studies show that "the key components of risk studies, such as probabilities, seem to have merely a minor impact in the media reporting" (Renn, 1991); other studies that look at media reports show quite the opposite. For instance, Freudenburg and colleagues after analysing 128 hazard events, they concluded that, "the amount of coverage is predicted only by the objective characteristics of the hazard events" (Freudenburg, Coijemb, Conzale, & Helgeland, 1996, p. 38).

Ultimately, the factors affecting media coverage of risk involve representations of diverse hazards. The identification of risk for decision making is valid when it is first approached in terms of the impact of hazards on risk. In decision theory the notion of risk can be expressed in the form of the "subjective expected utility" (SEU). The idea of subjective expected utility (SEU) suggests that for every hazardous event there is both a likelihood of the event happening, and a numerical utility that indicates the impact that he hazard would have on the target individual or organisation or a group of people. Decision-making theory defines risk as the product of the likelihood and utility. Hazardous events usually involve a number of different hazards and their summation gives rise to total hazard. The risk is, therefore, calculated as the sum of all the individual risks across all the events.

When the risk is expressed in the form of the SEU, individuals who are required to make a risk decisions in their everyday lives, first have to acknowledge a set of different possible courses of action, analyse those courses of actions into hazard events, and give values to the likelihood of every hazard and utility. They then calculate the risk of each course of action, add these risks and choose the course of action that gives the minimum total risk.

Making decisions in everyday life when expressing risk in the form of the SEU by analysing many micro-problems, gives rise to many obstacles related predominantly to the human nature limitations for employing subjective expected utility and making actual decisions in any literal way (Simon, 1990, p. 13; Simon, 1997). Another essential problem that contributes to the complexity of decision making is uncertainty about the reliability of what we know. We are not sure that we are aware of all the different parts of knowledge required about the set of different events that might give rise to the hazard.

The majority of individuals, when making decisions about their everyday lives, do not minimise total risk as in standard theory. Individuals often make risk-based decisions from a rational perspective in which the data are often personal. Lay-people's decisions about risk are predominantly dependent on personal interpretations of situations and decisions are taken from an intuitive and informal viewpoint. Nonetheless, as Slovic has argued: "lay-people sometimes lack certain information about hazards; however their basic conceptualization of risk is much richer than that of the experts and reflects legitimate concerns that are typically omitted from expert risk assessments" (Slovic, 2000, p. 191). The aim of this project is to understand more clearly the reasoning involved in making such risk-based decisions and how adults (35-63 years old) reason about risk in the big-data era.

The process of informal risk-based reasoning and risk interpretation is likely to call upon a coordination of a variety of different kinds of data involving explicit and implicit, quantitative and qualitative, and frequently value-laden.

Personal heuristics account for inconsistencies between an individual's judgment of a risk event (Brandstätter, Gingerenzer, & Hertwig, 2006) and the objective risk calculated by experts (Crossland, Bennett, Ellis, Gittus, & Godfrey, 1992). For example, availability—people's tendency to draw on knowledge that pops into mind quickly (Tversky & Kahneman, 1973; Kahneman, Slovic, & Tversky, 1982; Folkes, 1988)—and affect—emotional responses to highly unlikely events such as terrorist attacks on the airflight compared to most common events such as car accidents—both of which are shaped by personal experiences, contribute to cognitive heuristics for estimation of the likelihood of events are incorporated into individuals' judgment of risk (Greening, Gollinger, & Pitz, 1996).

In an effort to understand the complexity of informal risk-based reasoning, we acknowledge that real problems are extremely complex in their context-dependence, and they involve different levels of decision–making: individual (e.g., deciding about personal life insurance), social (e.g., deciding about the economic crisis in Europe), and mixed situations where there is an interplay between individual and social levels (e.g., in national child vaccination programs where parental choice is in interplay with official policy).

My interest lies in understanding how people call upon knowledge about the context throughout the process of risk-based reasoning and risk interpretation, and whether and how people coordinate likelihood and impact in their risk-based reasoning. I try to understand the processes used as individuals attempt to assess risk in various situations when presented with big data.

Purpose/Goals of the Study

In the study presented here, the project I set out underpins the investigation of adults' perceptions of risk in the big data era. In particular, going beyond the idea of risk in statistical theory, we attempt to gain an understanding of adults' perceptions and assessments of risk when they deal in our contemporary society with large scale data. The overall research study has involved working with a small group of adults 35-64 years old, recruited in pairs or individually from social media (e.g., Facebook, Twitter). This research study will be concerned with investigating adults' (35-64 years old) perceptions of risk in the big-data era and illustrating how risk is involved in decision-making in their everyday lives. An initial assumption is that decision-making involves the coordination of different kinds of information, based on quantitative models and personal value systems and judgements. To research decision-making when risk is involved, I develop decision-making scenarios (Appendix 1) for socio-economic, environmental, and health topics that involve modelling with personal value systems alongside using data from media or other (big) data sources. These scenarios will be used with adults' (35-64 years old) during an interview process.

In this research study, participants discussed various scenarios with the researcher, so that we can gather data on how people understand everyday situations that involve risk.

This research will also lead educators to a better understanding of adults' perceptions of risk when adults have to:

1. Make decisions about hazardous events; and,
2. measure the likelihood of events.

My research question, therefore, is: What are the personal values and models that affect individuals' thinking about risk and their process of reasoning.

Methodology

The total number of participants that was recruited was 40. The participants were initially contacted via social media and invited to participate in this research study. Selection was based on the potential participants who will first respond to the invitation.

The inclusion criteria required the participants to be aged 35-64, can use fractions and percentages to talk about risk and be aware about the dominant impact of risk in their everyday life.

In this paper we will discuss the results from the face – to– face interviews with four participants, each participant aged (35-44 years old). These four participants were mature students of a rural university who are studying for undergraduate degrees via online education. All participants were willing to collaborate with the researcher. The researcher sat with each participant or group of participants and asked participants questions about different scenarios for socio-economic, environment and health topics. Participants made formal or informal inferences while using information from Media or big data databases (sometimes Gapminder word map data). The discussions of the participants with the researcher were likely to be informal and as frank as possible. The researcher encouraged participants to explore and interrogate their own knowledge of the concept of risk in the big-data era and explain how risk figures in their everyday life. While these participants were keen to engage with the research study, the challenges that arose in interaction with the Gapminder world map raised questions about the software. The researcher intervened to demonstrate relevant aspects of the software, to address any technical issues, and to ask questions for clarification. The extensive discussions with the participants about diverse risk situations were recorded.

Camtasia video screen capture software recorded the reasoning process through the participants' extensive dialogues and manipulation of the Gapminder map. Data collected included audio recordings of participants' voices and video recordings of the screen output on the computer activity. The participants were asked to use the mouse systematically to point to objects on the screen when they reasoned about the various quantities presented in the Gapminder World Map graph. Having students to point on the screen helped to supplement the recording of their voices and explained their actions and interactions with the quantities illustrated by the Gapminder visualization tool that may otherwise have been subject to many interpretations. The recorded data were supplemented with notes and memos made by the researcher. The sessions lasted approximately 90 minutes.

The recordings were transcribed and analysed qualitatively. This process started with plain accounts for each participant and comparisons of the reasoning of each participant.

For the rest of this article we present the findings through the cases of two pairs: Tim (a student in environmental sustainability) and Michael (in Psychology); and George (in Sociology) and Steph (in primary Education). After presenting the chronological account of their engagement with the three scenarios, we discuss some key foci themes that emerged through the comparative approaches (Corbin & Strauss, 2007).

An Account of two pairs' activity

Tim (T) and Michael (M) began by trying to make sense of the information provided for the risk and heart attack scenario, interrogating the information and trying to assess the impact of taking medication to reduce the risk of a heart attack or stroke.

> *R*: Okay, and the second question 'Lupitor is a medication that could reduce of, by 36%. Will I buy Lupitor? Will I buy this medication, do you think it will help me or not?
>
> *M*: I think it will help yes.
>
> *R*: And how much will reduce the risk, of a heart attack?
>
> *T*: Umm, if we are just thinking just about the heart attack, I think it will reduce it by 36%
>
> *R*: Yeah, but have in mind that I had only 12% chance, risk of a heart attack so how much my chances will be?
>
> *T*: Well it will go …it will go from 12% to, 8 to 8and a half %
>
> *R*: How do you make this calculation?
>
> *T*: Well I just take 1, 36% is just over one third which is 33%
>
> *M*: So 1/3 of 12%, 1/3 of twelve is 4, so I take 4 off the twelve. Actually it will be more just under twelve because I'm increa…, I'm reducing it by more than 1/3. So it will be 7.8, but the 12% will go down to about 7.8 % or something around there.
>
> *R*: Okay, understand. So is it a good or does it reduce the risk by a lot or?
>
> *T*: I don't think 12% is a huge risk in any case… And I'll, maybe I'll think about not taking because I do not like taking drugs that much, so I would rather, try to do without the Lupitor.

Assessing the likelihood seemed to involve attempting to appreciate the relative size of probabilities. At this point, Tim and Michael felt that although the Lupitor could reduce this risk by 36%, the 12% risk of heart attack or a stroke would be reduced about 7.8%, and such a risk is not especially dangerous, so Tim came to a decision that he will not buy Lupitor.

In attempting to make sense of the information available as provided by the data taken from UK national mortality statistics, Tim began with interpreting the data that show the annual risk of death by age and sex in the UK.

> *T*: Okay, well, the first line "all ages", I think is everyone in the UK. And it says that in any particular year, one in every 136 men, and one in 193 women will die. And then it just breaks it down into different ages groups so if you're under 1 years old, you have, it's actually higher than I thought. 1 in 177 and 1 in 227, men and women respectively will die. The lowest age group that will, where people will die if 5-14. So 1 in 8333 men and 1 in 10417 women, well, at that age I should, 5-14 it's boys and girls will die. When you get to the 85 and over, it's because it goes slowly umm, the chances increase as you get older. So once you get to 85 and over, 1 in 6 men and 1 in 7 women will die. Umm but there seems to be a really big difference between 15 to 24.
>
> *R*: Why do you think?
>
> *T*: I guess that because of car accidents
>
> *M*: Drugs… and alcohol and cars and all those sort of things.

Participants attempted to make sense of the annual risk of death by age and sex in the UK for different ages, interrogating the information and connecting it with their everyday life experiences. While participants interpreted the annual risk of death by age and sex in the UK, appeared intuitively to try to explain risk attributed to different hazardous events.

Participants compared annual risk data of death by age and sex in the UK:

> *T*: In the age group of thirty-five to forty-four, one in every 1106 women will die. The men have higher chance of dying, 1 in 663 men. So, by the time we get to say, the number of women, it'd be almost 2 men would have died in the 35-to-44 age group.

Tim after comparing the risk data of death for females and males of the 35-to-44 age group in the UK, he found a relation between the two annual risk data. When the two participants, were asked to discuss the annual risk of death by age and sex in other countries, they were encouraged to use the Gapminder world map data.

> *T*: Ah, some countries had a life expectancy in some years which went down quite a bit. Other countries it went up, but not by much. The down side was quite a bit bigger but umm, I'm not sure I do understand. Do you mean where does the country, or does the world get a constant population?
>
> *R*: No I mean what is happening with regards to the risk of death in other counties.

The above excerpt portrays how the data presentation influenced participants' decision-making and caused tension between interpreting risks and focusing on the representation of variation derived from uncertainty as portrayed in the graphical representation of the Gapminder world map data. The animated Gapminder graph that shows variation and the change of the behaviour over time perplexed the participants who did not attempt to quantify of risk or uncertainty.

Then the researcher re-introduced the consideration for numerical quantification of risk:

> *M*: I've selected population growth, each year for the vertical axis and population total on the horizontal axis. It doesn't give an age and sex breakdown, we need some sort of measure of how many people are dying to get the risk of death. That's not the total, what happened? So,

crude death rate on the vertical axis and on the horizontal is still the population total (Figure 1).

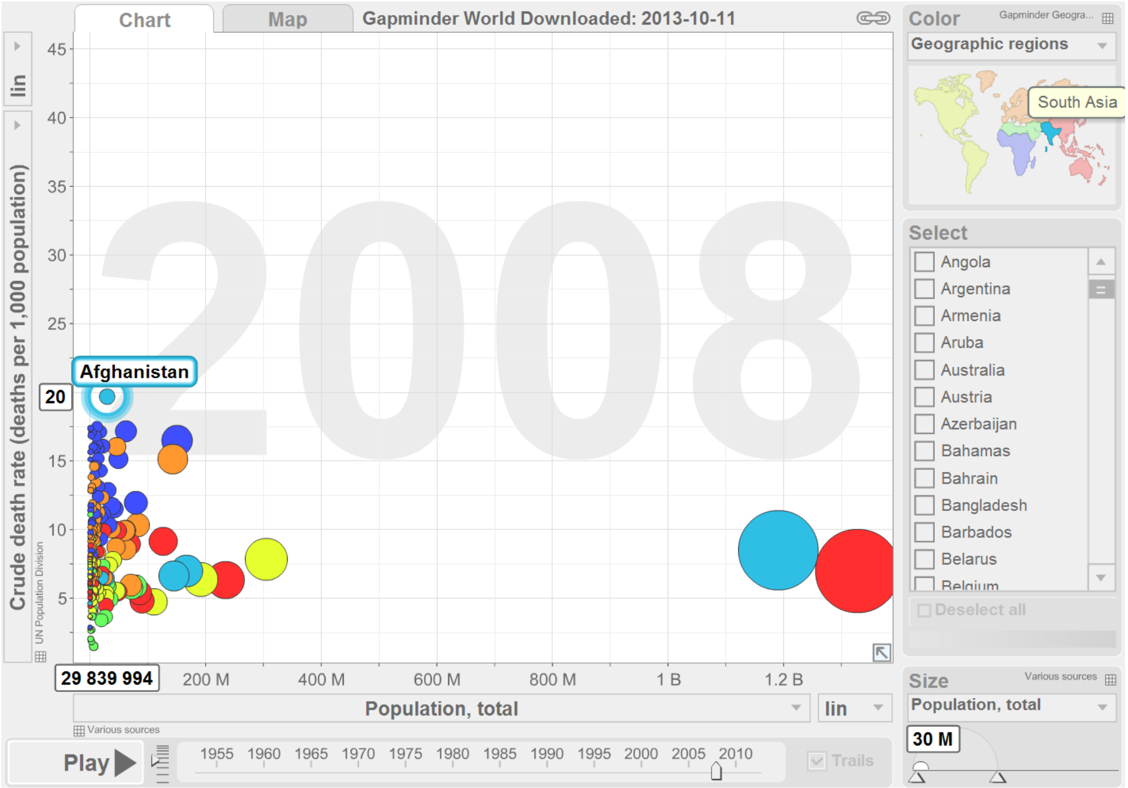


Figure 1. Crude death rate vs. Population total

R: What do you observe there?

T: Well India and China have the big populations in the world so they are on the far right of the axis, of the graph. Umm, and they have …… auh number of deaths per 1000 population

R: Can you see any percentage there?

M: 8.3%

T: So the total, so it's, for every 100 people three are 8.3 deaths

M: So 1 in how many, well 100 divided 8.3, ... , Or one say, divide by 8 is twelve and a half and one in twelve people

T: It sounds kind a high.

R: What about China?

T: 7.2, so one in fourteen. It doesn't sound right.

M: These two countries are the countries that they have the highest populations but not the highest death rate.

T: No, the highest death rate is Afghanistan.

R: Okay, which is how much

T: 19%

R: okay, what that means?

T: Oh no, I've made a mistake, I've made a mistake. It's nineteen deaths per 1000 population. So, one death every, it's close to twenty, twenty-two people or something like that.

R: How did you calculate it?

T: It's nineteen deaths every thousand people. So I divided the left hand side by nineteen and one thousand divided by nineteen. Oh no, I made a mistake because it's about fifty something. I need to get a calculator… So I'm dividing a thousand by nineteen, which is the deaths per one thousand population. And that gets me one death every fifty-two people.

The participants preferred to think in terms of actual numbers of people rather than probabilities or percentages. Often they had to take into consideration other social constructions (e.g., war) with which they were more familiar and interpret the data into their own schemata as shown below.

T: Well, if you're in Afghanistan, you have a much higher risk of dying. Umm, if you're in what was it, India you have fairly similar but you're a little bit, you're after in the United Kingdom than what you are in India.

R: Do you want to talk about Australia?

T: Seven.

R: Is it good?

T: Yes. Oh, one thousand, so in order to find out, three are seven deaths per one thousand people in Australia.

M: To find out how many people it takes for one person to die, I'd divide one thousand by seven…that is how many people are dying. Per one thousand, so it's one hundred and forty-two, umm, but that's both genders so, so it's probably safer if it's about fifty -fifty men and women, it's probably safer in Australia than what it is in the UK.

R: Why?

T: Because one in every one hundred and forty-two people are dying here in Australia. In the UK, it's something between fifty-fifty.

R: Okay, do you want to comment about something else, another country.

M: In the death rate, United Arab Emirates.

R: What is the risk of death by age and sex in the United Arab Emirates?

M: There, you've got umm, one and a half people …dying for every thousand.

R: Is it correct?

T: Yes, it's very, very few people are dying in the United Arab Emirate compared to the rest of the world.

The participants had to think of other countries, observed on Gapminder graph and the risk of death in those countries and compare these risks. The country that has the lowest death risk from the data representation on of data Gapminder attracted the attention of the participants.

When George (G) and Steph (S) engaged with the third scenario, which is about the risk of cancer due to alcohol consumption, they first looked at the two pictures that show the percent of alcohol-related cancer deaths per number of daily drinks consumed:

G: Percentage of high risk of cancer deaths based on daily number of drinks consumed.

S: So the high risk because we have 12% for women and 15% for men.

G: For men it's, umm it's high risk if you have 7 or more drinks per day. That doesn't seem to have a category on this on the bar chart.

S: Except to say more than 3 drinks per day. Uh so, well look… I think, I think umm, I think it's hard to say. Umm, well the questions ask "How do you understand the percentage of high risk of cancer death based on daily number of daily drinks consumed?" So, umm, this doesn't really look at risk in terms of high medium or low. Or high risk, risky or low risk the way the cancer council statistics provided so I think this doesn't even count.

G: The pie chart does not count ….well, it doesn't provide the information I need to answer the question that's provided. The other picture provides information (provided by the Cancer Council) that if you are a man and you have 7 or more drinks a day, then of 15% of men that do that have, will die in each year and for women if they have 5 or more drinks each day, 12% of them. Oh it really, it says are in risky, high risk category. I'm sort of making a bit of a leap that that means, you know, that they'll get cancer, they'll get cancer and not die. Umm 12% of women.

S: Well even that one doesn't provide that much information because what does risky, high risk mean? Does it mean that you'll get cancer? It doesn't say 20%, 12% of women will get cancer… it doesn't define what risk or high risk actually means.

In attempting to make sense of the information available, questions were raised about the actual meaning of the risk classes. It particularly concerned Steph that the high risk and low risk classes had no numbers associated with those classifications.

G: Actually I've just seen the rest of the graph, so you have 43% in this group in the la, for women, 43% … are in the low risk. So, it is 43% in the low risk, 12% in the high risk and 45% in the risky. But it still doesn't talk about what low risk, risky and high risk actually mean … I mean how many in those categories will get cancer. It all it says is that 43% are, have the habit of drinking up to 2 drink a day. And 12% have the habit of drinking of drinking 5 or more drinks a day. It doesn't say those that drink 5 or more drinks per day have a 50% chance of getting cancer.

S: All it says is that 12% of women drink 5 or more drinks per day and I'm starting to think about the pie chart as well. Umm, there different categories. One says up to 2 drinks a day, one says up to 1.5 drinks a day so, umm gain it's difficult to match the two, the pie chart to the glass graph…The more you drink, the higher the risk, it is that you will get cancer, but I don't know the numbers behind all of that.

Although the dialogue shows that the two participants are engaged with this scenario, it carries the implication that they cannot synthesise the information provided in the first graph that is the number of daily drinks consumed and the number of alcohol-related cancer deaths.

When they attempted to look at the graph estimated Alcohol consumption per average adult (15+) versus life expectancy on Gapminder map (Figure 2):

R: So the vertical axis will be alcohol consumption per average adult and the horizontal will be life expectancy.

G: Yeah because, umm well the, ah so you have Japan has the highest life expectancy.

S: What's this country, this has Moldova has the highest amount of alcohol consumption and the life expectancy.

G: Okay, South Korea. So …. Japan has the highest amount of alcohol consumption.

S: I think Japan has the lowest risk of cancer deaths.

R: Why do you think that?

G: Well because they live longer.

S: But what about the consumption of the alcohol.

G: Well it's in the middle. Umm…

R: What is happening in Afghanistan?

S: The life expectancy is low, probably because of conflict and war.

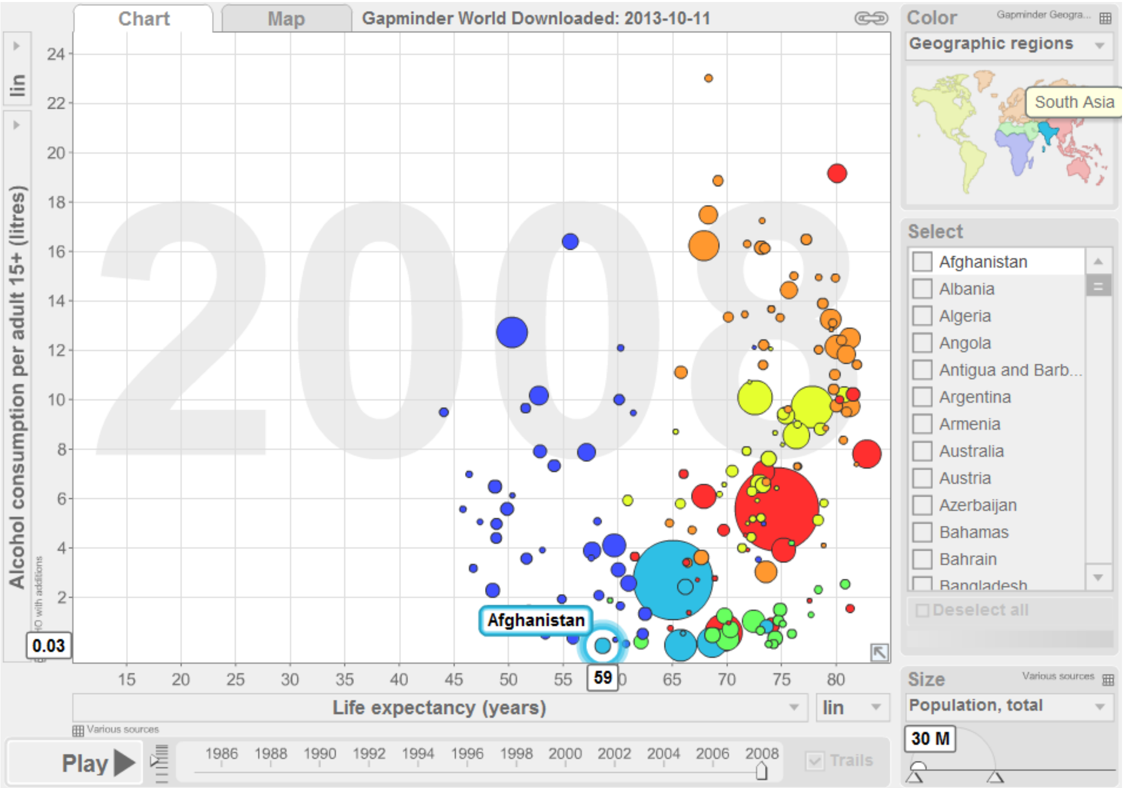


Figure 2. Alcohol consumption per average adult (15+) versus life expectancy

G: The alcohol consumption is very low, 0.03 litres for each adult over 15. Umm so, I'd say, yeah Afghanistan has the lowest risk of cancer deaths based on daily drinks consumed.

S: Ummm, and then I would say that Moldova … has the highest risk of cancer deaths because that's not a country in conflict so there's a lot of alcohol consumption. It's the highest alcohol consumption and it doesn't have a high, it's a medium but it doesn't have high life expectancy. So I think the consumption of alcohol that is occurring in Moldova is probably bringing the life expectancy down because of cancer.

George and Steph were more inclined to focus on the amount of alcohol consumption and the cancer deaths in countries where people consume the highest amount of alcohol as shown on the Gap minder map. When the relation of cancer deaths and the alcohol consumption are not linked, the participants attempted to coordinate alcohol consumption, life expectancy and qualitative data stemmed from the contextual politico-economic situation of the country. For example, they mentioned that the life expectancy in Afghanistan is low, probably because of the war. However, the participants commented that the alcohol consumption is very low, (0.03 litres for each adult over 15) and Afghanistan has the lowest risk of cancer deaths based on daily drinks consumed.

When Tim and Michael looked at the alcohol consumption and life expectancy in other countries, they commented:

T: Well it's interesting now that you say that, that the orange countries ... Europe and Central Asia …tend to have higher alcohol consumption and life expectancies than the African countries.

M: Yes, I see a pattern here

T: I'm going to umm, select GDP per capita. Alcohol consumption is per adult, GDP per capita, well most adults are generally the ones that contribute to the GDP. So … you could sort of say that umm, if income gets higher and higher, in-general you can say that alcohol consumption gets higher….And with that cancer due to caner due to alcohol consumption would be generally be higher but, obviously there's a couple of exceptions. Moldova again… they don't have a high income, but they have a high consumption of alcohol.

M: … but if a country like South Korea, they generally have a higher income and a higher consumption of alcohol.

T: That would be because of umm, Muslim bans on, on alcohol consumption so they have higher incomes but they don't consume much alcohol. So the likelihood is they probably don't have many cancer deaths because of… because they don't have a lot of high alcohol consumption.

M: I mean India and China are always interesting because they've got the biggest circles… India has a lower income and lower alcohol consumption, but China is, is higher income and alcohol consumption. But both are lower than Australia's….

When Tim and Michael introduce another variable that is the GDP per capita, they focus on alcohol consumption and GDP per capita and dismiss to refer to the risk data of cancer deaths because of alcohol consumption. Other features of the data visualisations of Gapminder, such as the size of the dots for China and India, indicating their large populations, attracted the attention of the participants.

Discussion

Despite the complexity of the risk-based reasoning, the participants were engaged with the three scenarios that required them to manage alongside the balancing of trading-off probabilistic evidence and a variety of different kinds data involving qualitative data inextricably derived with the context throughout the process of risk-based reasoning.

The responses of the participants largely became data that revealed how people are assessing and comparing risks. Almost all of the questions are what we might call "assessment questions" – they're questions asking for some sort of assessment or evaluation of a situation (e.g., "What is the income of the people of the countries that have higher and lower risk of cancer deaths based on daily number of daily drinks consumed?"). When the participants engaged with standard questions about risk involved in large-scale data sets, participants' reasoning was essentially based on mathematical calculations—for example the way that the participants worked through the calculation in the Lupitor example. They easily made inferences while reflecting on coordinating likelihood and impact in the context of their evidence.

When the participants were dealing with risk–based decisions about large-scale data, that required making sense of large collections of data, they "often achieve insight into big data by implementing a few principles that would make their data exploration easier:

•Explore the trends over time of fewer variables on the instantiated data structures;

•When faced with a trade-off in which there was clear best choice, they would create a new visualisation to further explore the question" (Prodromou, 2014, p. 71).

Using the interactive visualisations, they attempted to find meaning in the data, but in their visual justification efforts, they seemed to get distracted by other features of the data such as size of the population of a certain country or a pattern that the data seemed to follow, factors that add extra layers of complexity to the stuntedness and construction of risk.

In fact, features of big data appear to be highly salient factors that impinge on decision-making and frame the level of risk in situations that require consideration of large scale data. While the evidence from participants' dialogues suggests that participants made decisions based on salient features of the visual representations of large global datasets, quantification of risk was held back by issues around uncertainty and risk. Especially, participants acknowledged that they could not put numbers on risk when they dealt with uncertainty caused by large-scale data. In fact, participants made no effort to estimate risk using numerical values, or referring to any odds. They merely referred to risk using qualitative statements articulated in subjective terms based on their knowledge about the socioeconomic situation of the country of interest. Much of their discussions about risk were directed towards qualitative descriptors to provide rough quantification of risk.

References

Adams, J. (1995). *Risk*. London, UK: UCL.

BSON. (n.d.). The Notion and Definition of Risk. Retrieved from http://www.web-books.com/eLibrary/NC/B0/B60/006MB60.html

Brandstätter, E., Gingerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2), 409-432.

Corbin, J., & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd ed.)*. Thousand Oaks, CA: Sage.

Crossland, B., Bennett, P. A., Ellis, A. F., Gittus, J., Godfrey, P. S., Hambly, C., Kletz, T. A., & Lees, F. P. (1992). Estimating engineering risk. In F. Graham-Smith (Ed.), *Risk, analysis, perception, management*. London: The Royal Society.

Edwards, W., & Tversky, A. (Eds.) (1967). *Decision Making: Selected Readings*. Penguin Books.

Freudenburg, W., Coijemb, C. L., Conzale, J., & Helgeland, C. (1996). Media coverage of hazard events: analyzing assumptions. *Risk Analysis*, 16 (1), 31-32.

Folkes, V. S. (1988). The availability heuristic and perceived risk. *Journal of Consumer research*, 15(1), 13-23.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53-96.

Greening, L., Dollinger, S. J., & Pitz, G. (1996). Adolescents' perceived risk and personal experience with natural disasters: An evaluation of cognitive heuristics. *Acta Psychologica*, 91(1), 27-38.

Hilson**.** D. A**..** & Murray-Webster R. (2004). *Understanding and managing risk attitude: Proceedings of 7th Annual Risk Conference, held in London, UK*, 26 November 2004.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press.

Lupton, D. (1999). *Risk*. London: Routledge.

Pratt, D., Ainley, J., Kent, P., Yogui, C., & Kapadia, R. (2011). Role of Context in Risk-Based Reasoning. *Mathematical Thinking and Learning*. 13, 322-345.

Prodromou, T. (2014). Drawing inference from data visualisations. *International Journal of Secondary Education*, 2 (4), 66-72.

Renn, O. (1991). Risk communication and the social amplification of risk. In: R. Kasperson & P. Stallen (Eds) *Communicating Risks to the Public* (Dordrecht, The Netherlands, Kluwer Academic Publishers).

Simon, H. A. (1990). *Reason in Human Affairs*. Stanford, CA: Stanford University Press.

Simon, H. A. (1997). *Administrative Behavior: A study of decision-making processes in administrative organization* (Fourth ed.). New York: Free Press.

Slovic, P. (2000). *The perceptions of risk.* London, UK: Earthscan.

Spiegelhalter, D, (2012). Risk and uncertainty. Retrieved from http://www.cam.ac.uk/research/discussion/risk-and-uncertainty

Stirling, A. (1999). *On science and precaution in the management of technological risk (A synthesis report of case studies)*. Seville, Spain: Institute of Prospective Technological Studies (European commission Joint Research Centre). Retrieved from ftp://ftp.jrc.es/pub/EURdoc/eur19056en.pdf

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Weale, A. (Ed.). (2002). *Risk, democratic citizenship and public policy*. Oxford, UK: Oxford University Press/ British Academy.

Appendix

Questions

The interview questions will not be of a personal nature but general questions relating to the risk of death.

First Scenario: The risk of heart attack

Visit to my GP

•Recently I went to see my GP

•He told me I had a 12% chance of a heart attack or stroke in the next 10 years. What does this percentage mean?

•Lupitor could reduce this risk by 36%. Will I buy Lupitor?

Second Scenario: The Risk of death

Someone once said that the only certainties in this world were death and taxes. We look at tax rates with some interest, but give death rates much less attention, except when they are forced on us by some catastrophe, societal or personal. For that reason we thought it useful to have a reminder of the major effects on death rates - namely our sex, and our age.

Data were taken from UK national mortality statistics, which provides death rates per million population by age and sex. These have been recalculated to show the results as an annual risk - a chance of 1 in X of dying in the next year (on average), by your age and by your sex.

	Annual death risk 1 in X	
	Men	Women
All ages	136	193
Under 1	177	227
1- 4	4386	5376
5 - 14	8333	10417
15-24	1908	4132
25-34	1215	2488
35-44	663	1106
45-54	279	421
55-64	112	178
65-74	42	65
75-84	15	21
85 and over	6	7

Table 1: Annual risk of death by age and sex in the UK.

•Could you please explain to me the above numbers?

•What is the (percentage) risk of dying next year by your age?

•Which age and sex has the higher risk of death in the UK? Explain the percentages.

•What is happening in other countries? Please refer to Gapminder map.

Third Scenario: The risk of Cancer due to Alcohol Connection

I read online the following articles:

•The Cancer-Alcohol Connection, from the webpage:

http://www.everydayhealth.com/cancer/the-cancer-alcohol-connection-7199.aspx

and use the picture entitled:

"Alcohol-Related Cancer facts

• Cigs war won: Now cancer campaigners set their sights on beer, from the webpage: http://www.news.com.au/national/cigs-war-won-now-cancer-campaigners-set-their-sights-on-beer/story-e6frfkw9-1226088686962

and use the picture entitled:

"Australian Bureau of Statistics data showing the extent of problem drinking / Graphic: Vincent Vergara Source: news.com.au"

to answer the following questions:

•How do you understand the percentage of high risk of cancer death based on daily number of daily drinks consumed?

•Which country has the higher and lower risk of cancer deaths based on daily number of daily drinks consumed? Please refer to the graph estimated Alcohol consumption per average adult (15+) versus life expectancy on Gapminder map.

•What is the income of the people of the countries that have higher and lower risk of cancer deaths based on daily number of daily drinks consumed? Please refer to Gapminder map.

Calculated Risks:
The Teacher as Big Data Producer and Risk Analyst

Nat Banting
University of Alberta, Canada

Abstract: Teachers' work is often subjected to data analysis from outside sources in the forms of standardized examinations and media critique. This article uses the literature of risk analysis to play with two important analogies for teachers with regards to the emerging big data culture and the risk decisions therein. The complex context of the classroom facilitates the exploration of *teacher as big data producer*, while the multi-faceted nature of risk decisions provide the groundwork for the exploration of *teacher as risk analyst*. Illustrative classroom episodes portray examples of real and virtual risk faced by teachers, and a third category—curricular risk—is proposed.

Keywords: coherence, proportional reasoning, geometric representations, developmental perspective.

The availability of large data sets and development of complicated analytics is changing the decision-making processes in a multitude of disciplines. Purely data-driven approaches have become very successful in executing mechanistic processes, and offer valuable information in more complex, human-driven situations. (Spiegelhalter, 2014). The influence of big data and the decisions influenced by its analysis cannot be ignored in education. Teachers feel this impact in a number of ways, but none more immediate than third-party, large-scale examinations designed to satisfy an unending thirst for accountability and surveillance. While these examinations do collect data in a narrow sense (a detailed discussion of their validity is outside the scope of this discussion), they cannot compete with teachers as data producers and analysts of complex data. Here, I explore a shift of lens that frames teachers as the driving force in the collection and inquiry of big data rather than the inert consumers of third-party data. With a specific focus on mathematics education, my intent is to highlight two important roles of teachers in the emerging age of big data. A brief look at the existing literature on risk analysis highlights the importance of context and the existence of multiple rationalities in risk-based reasoning. This places teachers as the quintessential producer of big data and analyst of the pedagogical risks therein.

Teacher as Big Data Producer

The role of teacher is one steeped in context. If one were to frame the profession of teaching as a research study aimed at gathering data and testing various parameters, it would look much different than the data that institutions collect today. The multiple interacting forces of the classroom make it difficult to isolate variables and determine causality. Even when testing claims to decipher trends, they must do so in a contextual vacuum, because any deviance from the typical student in the typical classroom in the typical school (the generalization continues) is weakened by specificity. The result is a tendency to whitewash students into categories with those falling outside tidy lines labeled, rather fittingly, as 'at risk'. Large-scale testing, such as state or provincial math efficiency examinations, produces "large n, small p" data (Spiegelhalter, 2014, p. 264). The participants (n), although not always willing, are many while the parameters to be measured (p) are small in number. In the case of mathematics achievement tests, the participants might include an entire age group while the subject of the testing would include

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 378-388

isolated curricular concepts. The aim of such examinations is to produce a snapshot in time of the understanding of a specific subset of individuals. This snapshot is purposely contextless, and collapses the incredibly complex world in which teaching and learning occur. Student history, teacher style, school culture, and divisional priorities are necessarily ignored. The contextual void eludes to a quantitative objectivism in the subsequent risk decisions to be made based on the data.

The recent research on risk shows that risk decisions are based on a gamut of information and rationalities that involve tacit and implicit knowledge (Pratt, Ainley, Kent, Levinson, Yogui, & Kapadia, 2011). This stands in stark contrast to data collected with the purposeful eradication of context. In keeping with Spiegelhalter's (2014) definitions of statistical problems, education can be termed a "small *n*, large *p*" enterprise where teachers encounter small pockets of students (*n*) and collect vast amounts of tacit and explicit data (*p*) (p. 264). This data is linked to the context from which it emerges, but this is no longer considered a deficiency. Teachers encounter real problems that are "extremely complex in their context-dependence, and generally, dependent in reflexive ways on the subjective perceptions of different participant groups" (Pratt et al., 2011, p. 328). Teachers become the ideal producers of big data because they embrace the complexity of their task in context. Monteiro & Ainley (2007) point out that novice teachers often orient themselves toward practical answers, examples, and strategies when approaching a pedagogical situation. Their mindset is one of mastery through attainment, and mimics the neat picture painted by the typical textbook (Love & Pimm, 1996). It takes an experienced teacher to see that there are no straightforward answers, but only a constant reaction to the multitude of data sources being collected simultaneously. It is for this reason that experienced teachers often find themselves playing with textbook content, trying to fit it into the context of their practice (Love & Pimm, 1996).

I do not claim that teachers can possibly collect every ounce of incoming 'data' from a classroom experience; such an undertaking would be much too great to handle if we are to respect the intricacy of the task. Even the idea of 'collection' erroneously implies that teachers are exterior to the data. The teacher not only sees the context of the event, they participate within it. While test results view teachers as implementers or lay-people uninformed on the big picture of progress, Slovic (2000) shows that conceptualization of risk from groups outside of those considered experts include legitimate information about hazards that are typically omitted by the expert assessments. For this reason, a holistic stance on risk assessment in the classroom needs to include the intuitive viewpoint of the teacher, while the scientific approach is forced to narrow their scope and omit rich interpretations. Instruction becomes dynamic as formative assessment—the process of using dynamic classroom feedback to inform classroom risk decisions—collects data on various classroom risks (Wiliam, 2011). The teacher collects the data as they operate alongside students in the classroom milieu. It comes from numerous sources immersed in the context and history of the classroom. The teacher is uniquely positioned to collect data because of their participation in the classroom. The result is a feedback loop tailored to the immediate place and time—a teaching presence in the present. The dynamic nature of teaching creates a process of data collection that is equally fluid. This places the teacher as the optimum producer of big data and, in the absence of a hard, quantitative reality, the context becomes the reference for data production and subsequent risk analysis.

Teacher as Risk Analyst

Data production and risk analysis are not two separate processes, but inform and implicate upon one another. Classroom-based decisions involving risk are based on the contextual data produced by teachers, and provide data to inform future decisions. Risk is involved in any situation that involves uncertainty about the future outcomes. The uncertainty is categorized by two conditions: the likelihood of the hazard occurring and the impact should the hazard occur (Pratt et al., 2011). This traditional breakdown of risk assessment is useful for analyzing the teacher's role as risk analyst. Borovcnik and Kapadia (2011) argue that there are two levels of decision-making criteria: the personally preferred and the rationally bound. A personally preferred level involves a local analysis of the factors involved; it is a culmination of a personal perception and appetite for risk. It answers two questions: 'How risky is the situation?' and 'What risks am I willing to take?' A rationally bound criteria is a standard procedure that covers many different risk situations. Possible outcomes are assigned numerical coefficients of risk that measure likelihood and impact. The assignment of risk coefficients may vary depending on rationale. One rationality might fear events with greater impact but small likelihood of occurrence, while another may fear the sum of a series of likely events with relatively negligible impact. Institutions, including educational institutions, often favor the rationally bound, but the complex context of the classroom denies a single, defining rationality. There is no unique decision that leads to an optimal solution, because multiple stakeholders—teachers, students, parents, and administration—lead to many rationalities (Borovcnik & Kapadia, 2011).

Consider the difference in values between a school district vowing to raise mathematics achievement scores, and a classroom teacher trying to instill a deep understanding of mathematics in their students. The former would see any mistake in a negative context, while the latter might encourage students to voice mistakes and analyze their thinking. Differing value systems place the quantitative emphasis on different aspects of the risk analysis. The division wants to minimize the likelihood of a mistake occurring, because the negative impact on their mandate is large. This leads to them assigning a large coefficient to the impact of the hazard. A classroom teacher with aims at deep understanding views the impact of the event positively, and thus would implicitly assign a low negative risk factor. The risk coefficient would not be zero because they may fear a student's inability to self-analyze and move past mistakes. There is also the risk of said mistakes becoming entrenched. These differences in risk coefficients greatly affect how students encounter mathematics. Risk-based decision making from a single, rationally bound perspective is not reliable because multiple roles create multiple rationalities toward the assignment of risk-analysis factors.

Risk assessments must then involve more than a strict assignment of coefficients of likelihood and impact. Monteiro & Ainley (2007) argue that quantitative relationships must be balanced with previous knowledge, social context, and personal experience to achieve an accurate interpretation of a complex situation—a balance that the authors call *critical sense*. The teacher's context not only makes them the ideal producer of big data, but the most accurate analyst of the same data. Pratt (2011) calls the ability of a teacher to use intuitive statistical knowledge to reason with data *Informal Inferential Reasoning*. This once again validates the position of teacher as risk analyst because acting in a risky situation is no longer strictly a quantitative exploit, but "involves mobilizing a range of different kinds of knowledge and experience" (Monteiro & Ainley, 2007). The mathematics teacher has intimate access to classroom history, culture, and characters—all the ingredients of risk that Monteiro and Ainley define. They alone understand the nature of the risk and its mathematical roots; are trained in the psychological matters at hand; and know the context of the risk, information used, and people involved.

We can see these ingredients in play through the analysis of pedagogical tools created to study the teaching of risk. Teachers were asked to complete *Deborah's Dilemma*, a risk analysis simulation that used a mathematical model to simulate a patient's decision whether to undergo back surgery (Pratt, 2011; Pratt et al., 2011). Coefficients of risk were assigned to various outcomes with differing impacts ranging from full recovery to fatality. In the midst of a serious situation, the teachers quickly strayed from the quantitative relationships (although they always remained in play) and included personal lifestyle, past experience, and various other viewpoints in their arguments for or against the surgery. A particular teacher valued his athletic lifestyle. From his rationale, the benefits of regaining a sense of normalcy outweighed the chance of serious complication. One participant had recently experienced the death of a relative due to complications in major surgery. Their rationale included this past experience. Still another participant questioned the advice of the so-called expert in the simulation. They undervalued the doctor's recommendation because of the possibility of financial motives. The overarching theme became that the threat of serious hazard dominated over the likelihood that such an outcome would occur. Participants even changed their perspective of what constituted 'serious' or 'likely' as the pedagogical context unfurled. There was never consensus on a single, rationally bound criteria; each participant approached the complex situation with a critical sense. Their multi-faceted conception of the situation made them the ideal analysts of the risk therein.

Calculated Risks: The Risks Teachers Face

Mathematics teachers are encountering contexts much like *Deborah's Dilemma* on a daily basis. The medical context dealt with specialist opinions, past experiences, and various levels of physical impairment while the educational context deals with student histories, interrelationships, confidence, achievement, and time (to name a few). Teachers must collect data, analyze, and execute risky decisions. Risk analysis is not just another topic to include in an already lengthy curriculum; it permeates the activity of mathematics educators. Borovcnik and Kapadia (2011) define two types of risk situations both of which can be seen in a teacher's daily practice. A *real risk* situation involves a "severe impact already...and an action has to be taken to avoid more damage (p. 5505). A *virtual risk* is one of potentiality. No threat is currently posed, but there is conceivable threat for one to develop. I would add to these a third category of risk faced by teachers: *curricular risk*. Curricular risk situations occur *vertically* when classroom action provides the opportunity for the jumping ahead or backtracking through a topic of study, and *laterally* when classroom action encourages connection between several domains of mathematics that may not traditionally be linked in a linear curriculum. It is my proposal that the actions of data producing and risk analyzing teachers fall into these categories. For the purpose of illustration, four episodes accompany a brief discussion of the proposed categories. Although the exchanges are fictitious, they are created to model actions and conversations that typically occur within the domains of real, virtual, and curricular risk.

Episode 1: Real Risk

A grade five class has been going over the different ways to divide whole numbers. After many discussions and activities, the teacher is satisfied with the breadth of exploration on the topic. Before they are to move on, the standard long division algorithm—a mainstay in the curriculum and lightning rod for parents—needs to be covered. Knowing this, the teacher hopes that the previous mathematical experiences make for a smooth transition. After introduction of the algorithm, the teacher gives a series of questions in succession to see whether the students can apply the concept.

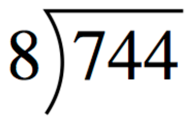


Figure 1. A typical long division exercise.

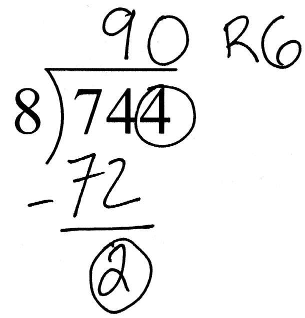


Figure 2. The student's work based on partial understanding of the algorithm.

One such question (Figure 1) asks the students to divide a three-digit whole number by eight. The teacher circulates the class and reassures the inevitable moments of hesitance from the students. One student has completed the question (Figure 2) and is sitting confidently in his desk. The teacher, noticing that the answer is incorrect, questions the student on their method.

Teacher: How did you get ninety?

Student: I asked myself if eight could go into seventy-four and it can. Nine times.

Teacher: Good, but you won't get exactly nine groups of eight.

Student: No. So I put a reminder of two because I have two left over. Then I asked myself if eight could go into four, but it can't.

Teacher: Where did you get the four from?

Student: I dropped it down from the top. Eight can't divide it so that's zero times and another remainder of four.

Teacher: What do the circles mean?

Student: I circled the remainders. Two and four make six. Ninety remainder six.

The student in this episode has arrived at an incorrect solution, but has provided a wide array of data. They have clearly taken some of the algorithm's rhetoric to heart. The ideas of grouping, dropping down, and remainder are present but malformed. Handling mathematical errors is the most recognizable type of real risk in the mathematics classroom. At first glance, this problem might be addressed with a re-explanation of the method and more practice with the steps, but the amount of detail provided by the student coupled with the teacher's unique understanding of the classroom and student creates a "small *n*, large *p*" risk analysis. I do not mean to villainize a review of the algorithm, because it may present itself as the best choice after consideration of the data. Time, learning style, and other student needs could all point to reiteration as an effective plan of action. The danger of such responses is they become insensitive to the data and part of an automatic instrumentalist approach of transmission and, if necessary, re-transmission.

The student's incorrect response presents a real risk to their understanding of the topic of division. The presentation of an algorithm has bankrupted much of the other conceptualizations of division previously addressed in class. In this light, classroom situations involving real risks often jeopardize much more than the problem at hand. They can deteriorate a solid mathematical foundation. The necessary actions cannot be predetermined in lesson plans; they are created in direct response to classroom data production and the subsequent analysis of the risks involved. The teacher becomes a calculated professional operating from their unique character; the response of the teacher reveals their rationality—what they value. Handling real risks is a process of constant recalibration based on the misconceptions of an ever evolving classroom consciousness.

Episode 2: Virtual Risk

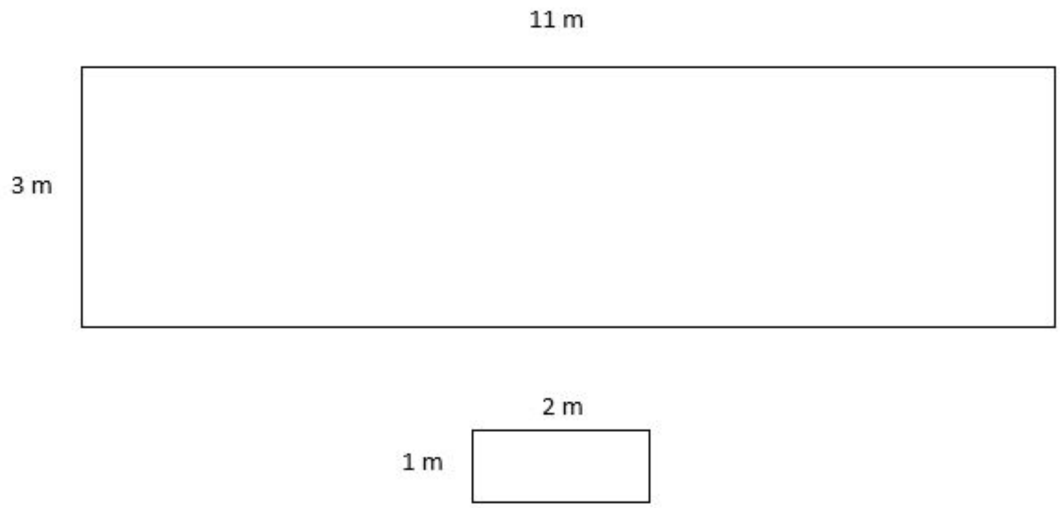


Figure 3. A diagram depicting the mural task.

Grade ten students in an applied stream of mathematics have chosen partners to complete a series of small tasks involving measurement. A few students have chosen to work on the projects alone. The teacher has allotted a three-day chunk for the completion of four tasks with varying levels of difficulty. The hope is that the set of tasks will provide some data not only on the mathematical abilities of the students, but their work habits as well. Both of these factors are crucial when beginning a new semester of coursework. One of the tasks asks the students to hang a rectangular mural exactly in the center of a large, rectangular art gallery wall. The mural measures one meter by two meters, and the wall is three meters high and eleven meters long (Figure 3). As the class works, several strategies emerge; the teacher takes mental notes for a discussion before the end of the class. One student working alone raises their hand and asks a familiar question.

Student: Is this right?

Teacher: What did you get?

Student: I figured out that the mural should be hung with one meter of space top and bottom.

Teacher: Your diagram looks great! How did you get one meter?

Student: The wall was three meters tall, and I needed space for above, below, and the actual mural. So I divided the wall into three even portions.

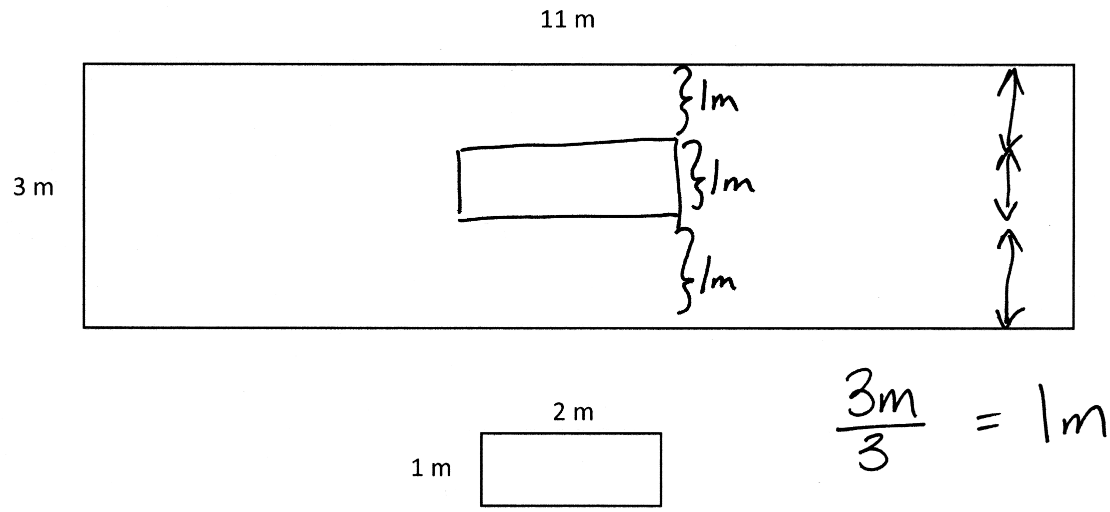


Figure 4. Diagram of the student's work on the mural task.

The student's work (Figure 4) has pieces of truth littered throughout it. It addresses the idea of division as an operation that splits objects into equal pieces. It also shows that the student can connect the abstract operation of division with the concrete process of measurement. The situation is one of virtual risk because the over generalization creates potential for future complications—instances of real risk. The teacher quickly realizes that the answer is correct, but the method is flawed. If the student were to execute the same process to complete the next section of the project, they would arrive at an incorrect solution. The potential for errors makes virtual risk difficult to diagnose and, by extension, difficult to handle.

Virtual risk eliminates the reiteration of method as a possible solution because the method, built on partial or overgeneralized truths, arrived at a correct solution. A popular response is to prompt students to apply their method under different circumstances. In this case, the teacher could ask the student to complete the project by working out the necessary distance to the left and to the right of the mural. If we view teaching through the lens of risk analysis, the situation may call for an entire-class discussion of the method and its limitations. Such a decision would need to be based on the context of the classroom. Would the learner be embarrassed? Have others made similar mistakes? In general, will encountering the thinking impact the class positively or negatively? What are the risks and what are their impacts?

Virtual risk separates itself from real risk in the sense that curriculum planners often entertain situations of virtual risk because they deem them too complex for the students at a current stage of mathematical development. The use of language is crucial to these creations. A textbook—and teacher—in primary school might assert that you cannot subtract a larger number from a smaller number. This is, of course, not true, but deemed necessary at the time. This decision is made knowing the risk of students cementing this notion. Similar instances occur when teachers state that you cannot take the square root of a negative number, or that zero divided by *anything* is zero. The language of partial truth creates conceivable, virtual risks in the student's future. The willingness of the teacher to dwell in, and debate the merits of, virtual risk reveals their rationality once again.

Episode 3: Vertical Curricular Risk

In an attempt to have students harness their human tendency to categorize, the teacher has students in a grade twelve class divide into groups of three and work together to solve a problem in set theory without any formal introduction to its language, notations, or formulae. Students work in their think-tanks as the teacher circulates to provoke thought with leading questions. One particular group's work looks scattered from a distance, so the teacher goes to investigate. The group has used a primitive graphical organizer to solve the problem (Figure 5), but because the teacher does not immediately recognize it through the mess, a conversation ensues.

A particular school has 56 Grade 12 students. Of these, 21 play football and 19 play basketball. How many students play both sports if there are 25 Grade 12s that do not play either sport?

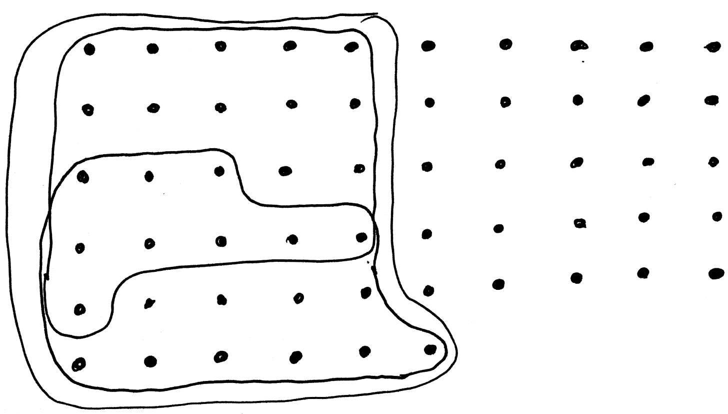


Figure 5. The group's visual solution to the set theory problem.

Teacher: Can someone explain what all of this means?

Student 1: Each dot is a student, and each line is a group.

Teacher: Okay. So where do these belong? (Pointing to the dots not circled)

Student 1: Those are the kids not on either team.

Student 2: Think of them like the left-overs.

Student 1: Then we drew the line around the football players, and a separate line around the basketball players.

Teacher: How did you know where to start?

Student 2: We didn't, but knew every one of these (pointing to the circled dots) needed to be on a team

Student 3: Or they would be in the left-overs. But they can't, because the problem says so.

Teacher: Okay, looks nice. How did you solve the problem then.

Student 1: By counting. (Pointing to the intersection of the two sets).

The students created an inefficient—but personally relevant—version of a Venn diagram. The teacher was not planning on introducing this specific curricular outcome until students got the implicit hold of the various concepts to be mastered—union, intersection, and universal set. The teacher is faced with a curricular risk, a decision that affects how the students will encounter the curriculum content. In the case of vertical curricular risk, the students have already formulated an idea of what is to come; it is evident that the group would benefit from the introduction of a formal Venn diagram. The risk is intimately tied to context because the teacher must make the decision if the rest of the class is ready to jump forward.

The language 'jump forward' implies a linear teaching model, but I am not implying that learning takes place strictly in this way. I am simply mirroring the incremental nature of most curriculum guides and textbooks in mathematics. Whether or not the new skill is further along the continuum of abstraction or sophistication is up for debate. Even the debates of whether this continuum exists or if abstraction is the ultimate goal of mathematics are well beyond the scope of this discussion. Vertical curricular risk is encountered when the anticipated class structure is perturbed by student organization. Episodes like the one above are far from negative. They differ from real risk in the sense that no mistake has been made; they cannot be classified as virtual risks because no potential for future threat exists. The risk is one of readiness. Vertical curricular risks provide teachers yet another opportunity to operate as risk analysts. They must gather data on the curriculum, the students, and classroom culture and decide whether the vertical risk is one worth following, or banking for use at a later date.

Episode 4: Lateral Curricular Risk

We move back into the classroom from the second episode where grade ten students are working on the mural task (Figure 3). A pair of students has worked quietly for upwards of fifteen minutes but become focal in the teacher's attention when they begin a hushed argument. After a short time to allow the debate to continue, the teacher moves to their table to ask what the argument is about. The students have very little written on their page, only a single diagram (Figure 6). Initially, the teacher diagnosed the problem as a group off task (a type of real risk), but conversation revealed a much different problem.

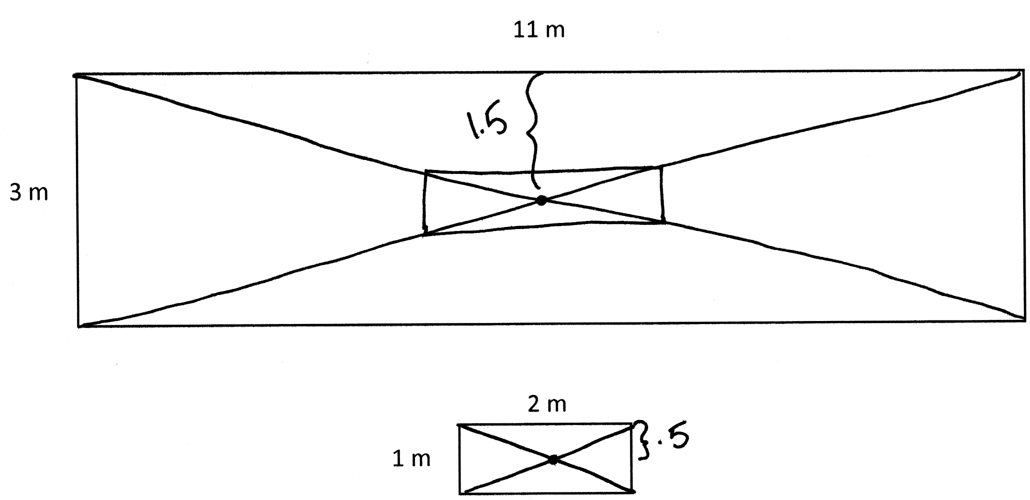


Figure 6. Diagram of the students' geometric solution to the mural task.

Teacher: What's the problem here?

Student 1: I think we solved the problem, but is it bad that we solved it without any math?

Teacher: What do you mean?

Student 2: See, I told you it was wrong. We didn't measure anything.

Student 1: We got the answer. All we did was connect the corners on the wall and on the mural. They both meet in the middle.

Student 2: He wants us to match up the two middles to hang it in the center.

Teacher: Would that work?

Student 1: If you put the two centers together, it must be in the center.

Student 2: But how do we know that we actually found the center? We are just guessing.

This episode illustrates the plurality of risks that a teacher encounters and further supports the analogy of teacher as big data producer and risk analyst. Each interaction with student thought builds on the cache of data that can be used in ensuing risk decisions. Risk decisions become interconnected as the data becomes entrenched in the activities of the class. Here, the teacher encounters a particular orientation to mathematics. For these students, mathematics is perceived as a process of numerical manipulation, not a process of argument. The crossing of the rectangles, although a productive geometric argument, did not contain the structured actions typical of the applied strand of mathematics. The teacher, knowing this history, is then faced with a risk decision to introduce the topics of symmetry and geometry, or reroute them back to the intended outcomes. Implicated in the decision are the teacher's impressions of mathematics (Davis, 1995). All classroom data is filtered through the teacher as risk analyst, so if they believe school mathematics is to be connected, they may be more likely to take the shift presented in the lateral curricular risk. The curricular risk here is lateral because the question is not whether to move forward or backward within a line of curricular study, but whether to bridge the gap between two topics, the topics of measurement and geometry. Lateral curricular risks are the most seldom executed because of the dominant rationality toward a linear curriculum.

Conclusion

This discussion situated classroom activity in the culture of big data in order to attempt to shift the role of teacher away from that of passive decision maker and into one of active risk analyst. The intricate and contextual data collected by teachers resembles that of the big data culture affecting many disciplines—only some of which are highlighted by Spiegelhalter (2014). Teachers are uniquely situated in their classrooms to take on the role of big data producer because of the intimate familiarity of the context to the point of co-implication in its construction. While there are a multitude of rationalities to approach each decision, teachers are best able to balance the various factors from stakeholders, filter through their own impressions of curricular mathematics, and execute risky decisions. The categorizations of risk—expanded upon from Borovcnik and Kapadia (2011)—are not attempts to simplify the muddy process of risk analysis; such a rationale is what we aim to move away from in the first place. Rather, they highlight that teaching is not a craft of predetermined reactions along a single path of best efficiency, but a constant recalibration of data collection and risk analysis.

References

Borovcnik, M., & Kapadia, R. (2011). Determinants of decision-making in risky situations. *International Statistical Institute: Proceedings of the 58th World Statistical Congress*, *Dublin*, 5503-5508.

Davis, B. (1995). Why teach mathematics? Mathematics education and enactivist theory. *For the Learning of Mathematics*, *15*(2), 2-9.

Love, E., & Pimm, D. (1996). 'This is so': A text on texts. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education* (pp. 371-409). Dordrecht, The Netherlands: Kluwer.

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education, 2*(3), 187-207.

Pratt, D. (2011). Re-connecting probability and reasoning about data in secondary school teaching. *International Statistical Institute: Proceedings of the 58th World Statistical Congress*, *Dublin*, 890-899.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning, 13*(4), 322-345.

Slovic, P. (2000). *The perceptions of risk*. London, UK: Earthscan.

Spiegelhalter, D. J. (2014). The future lies in uncertainty. *Science 345*(264). 264-265.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Dividing a pizza into equal parts – an easy job?

Hans Humenberger[1]
University of Vienna, Austria

Abstract *Theoretically seen dividing a pizza equally is not an easy task. For instance, with a normal knife (straight cuts) one has to hit the center so that the cut is a diameter. But there are alternatives (also for dividing equally between more than two persons) which have strong connections to elementary geometry and to integral calculus. This paper deals with these alternatives elucidating the so called "pizza theorem".*

Strictly speaking dividing a pizza into equal parts is not as easy as it may seem at first glance. Even if it is to be shared only between two people and the pizza is circularly shaped. After all, one has to hit the center so that the cutting line is the precise diameter. Cutting the pizza into roughly equal pieces will not be a problem at all in real life. There will normally be no conflict over who gets which piece. But what if the pieces are to be completely exact? Of course, such considerations are more theoretical than practical in nature, but they may provide useful mathematical and didactical input for teaching mathematics at different levels. In fact, in mathematics important questions are not always practical, but in some cases only theoretical.

1 The phenomenon of equally dividing a pizza ("pizza theorem")

There is a possibility using a pizza knife consisting of four straight blades with "center" *P* (*P* divides every blade into two parts, adjacent blades always have an angle of 45°) to carry out in reality a division of a pizza that is also theoretically exact. One can imagine this knife as a special cutter (the center *P* is put anywhere on the pizza) pressed with power onto the circular pizza so that afterwards there will be eight pieces (Fig. 1). The shape of these pieces is very similar to sectors of a circle but they are not really ones (except in the case that *P* hits the center of the circle), however we will call them "sectors" for simplicity reasons.

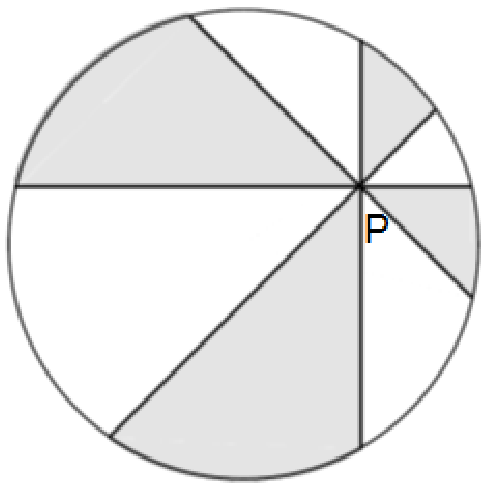


Fig. 1: Cutter on a pizza – schematically

If the first person takes every second "sector" (e. g. the white ones in Fig. 1) and the second person the remaining ones (grey in the figure) then the pizza has been equally divided! A realization of the division process with a real pizza and a specially prepared "pizza knife" (cutter) can be seen in Fig. 2.

[1] hans.humenberbger@univie.ac.at

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 389 – 403

Fig. 2a: Cutter on a pizza – in reality

Fig. 2b: Pizza after division

This is probably surprising and does not fit in some sense to the symmetry conditions of the circle, hardly anybody would guess this intuitively. On the contrary, formulated as a question most people would negate that this procedure leads to a really equal (also theoretically) division. Nowadays with DGS ("Dynamic Geometry Software") one can see the corresponding phenomena: Every DGS can measure areas and add up these measures (the areas are not real sectors but they can be separated in triangles and circular segments and these types of areas can easily be measured). One could produce a corresponding applet that works in that way (when moving P the area sums of the grey and white "sectors" respectively remain unchanged). Using hands on methods one can see this by cutting a circular piece of carton in the described way and weighing all the white and grey pieces at a time.

Remark: Kroll/Jäger 2010 propose to formulate this problem for teaching purposes as a story of distribution of an estate. I don't think that this is a good idea because this formulation leads away from the mathematical problem and sounds rather artificially (treetops as circles etc.). On page 101 one can read (translated): "Nobody should be bothered by the fact that the problem is not realistic; students are not either." Even if this should be correct (I really doubt that!) I consider this as quite problematic, formulations like this can easily lead to wrong and in some way "dangerous" beliefs about applied mathematics.

The problem of finding corresponding proofs – meant as a problem to be solved individually by students at school – is surely too difficult. But nevertheless this phenomenon has something fascinating and motivating to deal with it more intensively. What could a concrete teaching unit look like in which this phenomenon is dealt with? In which grade could this be done? Below we give answers to these questions.

Firstly we give mathematical analyses of our topic in two ways, on the one hand using elementary geometry and calculus on the other. Hereby we also deal with some references.

2 Elementary geometry

2.1 Presentations following S. Wagon and others

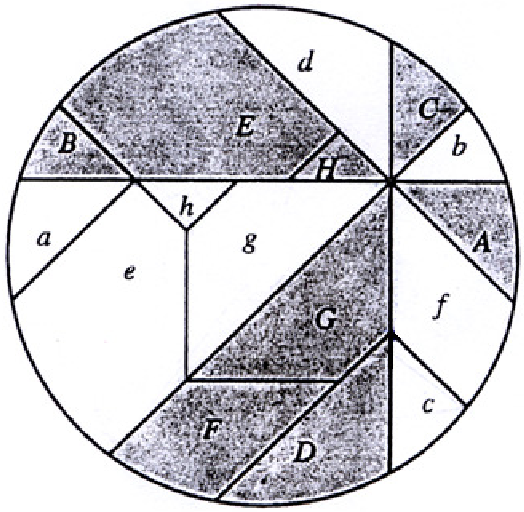


Fig. 3: Dissection following Carter and Wagon 1994

The following reference is somehow striking (at least if one is not familiar with this phenomenon). In the very short article of Carter/Wagon 1994 (1/2 page!) one can see a "proof without words", just a picture (Fig. 3), no explanations.

This picture shows (autonomously one would hardly come to similar results) that there seems to be not only equality of area sums but also that the grey and white areas can be dissected into pairwise congruent figures (equidecomposability). That means using the same "puzzle pieces" one can build up the white and the grey area. This is even more surprising than the mere equality of the area sums. The above mentioned word "striking" refers to two aspects: (1) The equidecomposability, in other words the possibility of puzzle pieces. (2) Is the corresponding proof really so easy that it is not necessary to use a single word of explanation?

Using corresponding letters (capital letters and small ones) for the areas Carter/Wagon indicate that the corresponding areas are congruent. But interpreting the picture itself (ideas behind, thoughts, reasoning, etc.) is not so easy. How can we find words? What happens in Fig. 3? How does the dissection emerge, how is the figure built up (construction)? Which lines are presumed to be parallel or equally long? Which angles are presumed? In the end: *why* are areas with corresponding letters congruent? The congruencies $A \cong a, B \cong b, C \cong c, D \cong d$ are easy to explain: In these cases probably a reflection took place on the horizontal and vertical diameter (not drawn in the figure); g is the reflection of G (with the common border as the axis of symmetry). But why do the other congruencies hold $(E \cong e, F \cong f, H \cong h)$? Relations that need not to be communicated in a mathematics journal ("proof without words") are from another point of view not self-evident. There is still a wide range of possible interpretations, how can the structure of reasoning for the congruencies be built up? The picture itself does not say how the authors thought, how the figure arose in their mind. Before I got to know other figures concerning our topic (see below) I gave this problem to student teachers in a geometry course: analyzing the above "proof without words". Some of the students came to important (partial) results but nobody could give a really correct and consistent analysis.

Such an analysis can be given in several ways (focusing either on the idea of congruence or on geometric transformations) but it is not easy for student teachers to establish detailed and precise reasoning. Viewers of Fig. 3 mostly don't see arguments for the congruencies $E \cong e, F \cong f, H \cong h$ at a glance, so the title "proof without words" may seem to be not so appropriate (even for mathematics students at university).

A very similar version of the above Carter/Wagon dissection is given in Kohnhauser/Velleman/Wagon 1996, p. 118 (Fig. 4), the pieces h and H are half as big as in Fig. 3. But the crucial difference to Fig. 3 is given by the fact that in Fig. 4 there is a highly symmetric octagon *PQRSTUV* indicated which is very helpful for establishing a possible proof. The "genesis" of this octagon can be thought like this: The rectangle *PSTW* (its center is the center *M* of the circle) is rotated by 90° (center *M*) and the rotated points *Q*, *R*, *U*, *V* must lie on the "45° lines" through *P* and *W* (why?). The octagon *PQRSTUVW* is mapped onto itself under the 90° rotation with center *M*, and this can be a crucial hint when looking for some reasons for the mentioned congruencies. In my opinion it is more justified to call Fig. 4 a "proof without words" than Fig. 3 (here the octagon is missing and therefore viewers have no hint concerning the important 90° rotational symmetry).

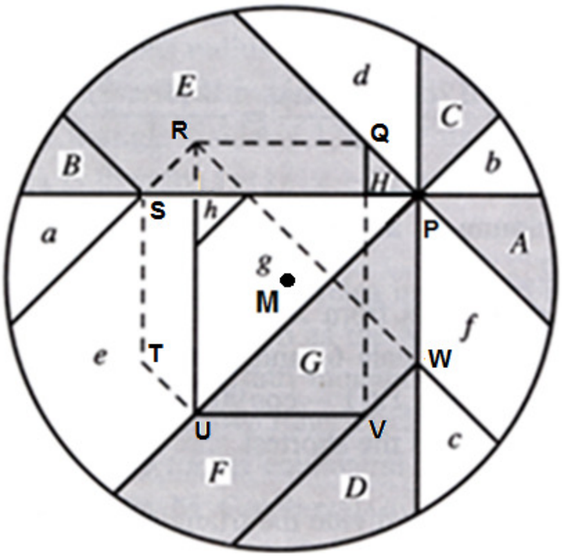


Fig. 4: Dissection – including an octagon

The congruencies $A \cong a, B \cong b, C \cong c, D \cong d$, $G \cong g$ can be explained in a similar way as above (reflections). For the other congruencies one could argue using the highly symmetric octagon as follows: Because of the octagon symmetry (e. g. the sloped diagonals are of course "45° lines") we can reason for $H \cong h$ (isosceles and rectangular triangles with the shorter octagon side as hypotenuse; in Fig. 3 the congruence $H \cong h$ is also apparent but how should one prove it?) The congruence $F \cong f$ could be proved like this: When rotating the octagon and *f* by –90° (rotation center *M*) the octagon maps onto itself and *f* maps onto *F* (why?), and therefore *f* and *F* are congruent (analogous: $E \cong e$). Due to the octagon one probably sees more easily (and can give reasons for it) that $f \mapsto F$ and $e \mapsto E$ under the mentioned rotation. Using Fig. 4 I have no experiences up to now how successful mathematics students are in explaining this "proof without words" but I suppose the success rate is higher than using Fig. 3.

2.2 Presentations following P. Gallin

In Gallin 2011, p. 12 one can find a striking and very simple proof in which very many parts are needed and therefore this proof may seem a bit confusing at first glance. But one should not be "scared" by the huge number of pieces although it may take a while to fully understand the simplicity and brilliancy of the ideas behind. The corresponding figure (Fig. 5) could really be a "proof without words" even from the perspective of students, e.g. with the following text: *P* and the four "blades" are reflected on the point *M*, on the two coordinate axes, and on the angle bisectors of the coordinate system so that the reflected points yield an octagon (drawn thickly[2]). Now the outer and the inner part of this octagon may be considered separately and the equality of the area sums of the colored and white pieces respectively can be seen almost directly (even the dissection into pairwise congruent pieces).

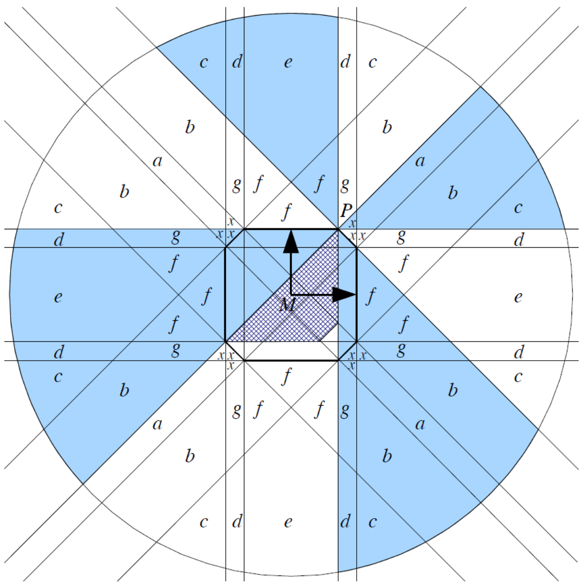


Fig. 5: Dissection following P. Gallin

[2] Due to these reflections a very high degree of symmetry is established – this we have neither in the initial situation nor in Fig. 3 and 4.

"Solution": Outer parts: Due to symmetry reasons (these need not be explained further in this situation) one finds congruent pieces in the colored and in the white area: In each area 2 pieces *a* and *e*; 4 pieces *b*, *c*, *d*, and *g*; 6 pieces *f* and *x*. Hence outside of the octagon the area equality is clear. Inner parts: The hatched trapezoid belongs to the white area and is congruent to the adjacent colored one and also the other parts within the octagon are easily recognized as consisting of congruent pieces (one mall trapezoid and four small triangles).

The octagon of Fig. 5 is in particular the same as in Fig. 4. The ingenious idea of P. Gallin is to do the reflections not only with *P* but also with the four blades of the cutter. This leads to a highly symmetric configuration in a situation that is in its origin (Fig. 1) not symmetric at all, Gallin has reached the highest possible level of symmetry by his method. The prize for that is a huge number of lines and area pieces but one can interpret the resulting figure in an easy way due to its symmetry. The congruencies of all the area pieces *a* (or *b* etc.) become completely evident, further reasoning for them is not needed. This feeling one does not have looking on Fig. 3 (and still in a weaker form with Fig. 4).

In the paper of P. Gallin (2011, p. 14f) there is also another proof (with elementary geometry) for the phenomenon of the equal area sums.

By the way, not only the area is divided equally also the boundary of the pizza, nobody needs to eat more of the perhaps pretty dry boundary which is often not so in favor because it tastes sometimes like ordinary bread. Also for this phenomenon (pizza boundary, arc length) there is a elementary proof (cf. Gallin 2011, p. 13f) that the sum of the four arc lengths of the colored pieces is equal with the corresponding sum of the white ones.

One can easily show that – having chords with a constant angle α – the sum of the arc lengths is also constant (in particular independent of the position of the intersection point *P* and the special position of the chords). The sum of the arc lengths only depends on the angle α (cf. Fig. 6a): $\left|\widehat{A_1C_1}\right| + \left|\widehat{B_1D_1}\right| = \left|\widehat{A_2C_2}\right| + \left|\widehat{B_2D_2}\right|$.

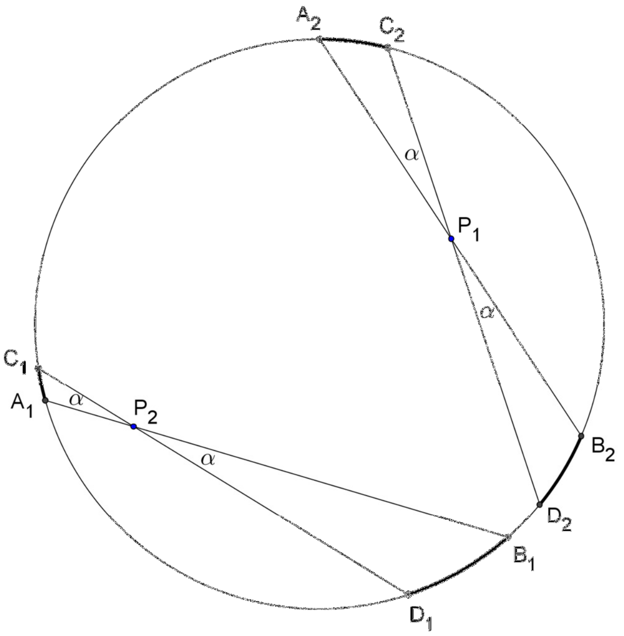


Fig. 6a: Equal sum of arc lengths

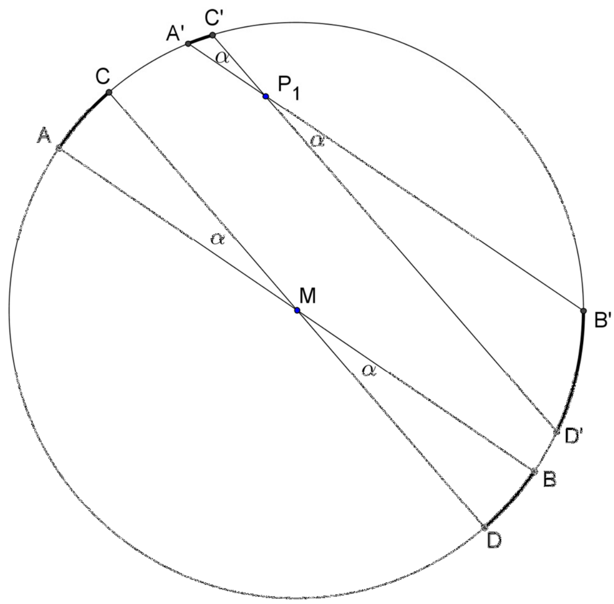


Fig. 6b: Pairwise parallel chords

A crucial idea is here: It suffices to show this for $P_2 = M$ and pairwise parallel chords (see Fig. 6b): $|\overset{\frown}{A'C'}| + |\overset{\frown}{B'D'}| = |\overset{\frown}{AC}| + |\overset{\frown}{BD}|$. It will turn out that the lengthening and the shortening in the transformations $|\overset{\frown}{AC}| \to |\overset{\frown}{A'C'}|$ and $|\overset{\frown}{BD}| \to |\overset{\frown}{B'D'}|$ are cancelling each other, and therefore in the end we have no change in the sum of the arc lengths. We have:

$$|\overset{\frown}{A'C'}| = |\overset{\frown}{AC}| + |\overset{\frown}{CC'}| - |\overset{\frown}{AA'}|$$
$$|\overset{\frown}{B'D'}| = |\overset{\frown}{BD}| + |\overset{\frown}{BB'}| - |\overset{\frown}{DD'}|$$

and because of $|\overset{\frown}{AA'}| = |\overset{\frown}{BB'}|$ and $|\overset{\frown}{CC'}| = |\overset{\frown}{DD'}|$ (this is clear due to the symmetry of the circle) we get by summation: $|\overset{\frown}{A'C'}| + |\overset{\frown}{B'D'}| = |\overset{\frown}{AC}| + |\overset{\frown}{BD}|$. Also the value of the sum is easy to see in the case that both chords pass through the center *M*, the value is $2\alpha r$ (α in radian measure).

Remarks:

- Considering the *area* one needs for constancy *all four* parts (sectors), in other words two "double sectors"; considering the *arc lengths* one needs for constancy only *two sectors* (one "double sector").
- An alternative way of reasoning we give below with calculus.

2.3 An important lemma of elementary geometry and a plausibility consideration

In the focus of the considerations above we had the phenomenon that the four colored "sectors" and the four white ones can be divided into pairwise congruent pieces (equidecomposability), this is more than the mere equality of the area measures. In the following we concentrate on the equality of the area measures using a completely different approach which is building a bridge to the next chapter (calculus).

Lemma: For every orthogonal pair of chords in a circle (segments *a*, *b*, *c*, *d* – see Fig: 7) following equation holds (*r* is the radius of the circle): $a^2 + b^2 + c^2 + d^2 = (2r)^2$

For this lemma there are several possibilities for proving it, a typical case for *problem solving*. A short and illustrative "proof without words" is given in Nelsen 2004.

This lemma plays an important role on the one hand in the following argumentation with *plausibility* (not a rigid proof) and on the other when we use methods of *integral calculus* (see below).

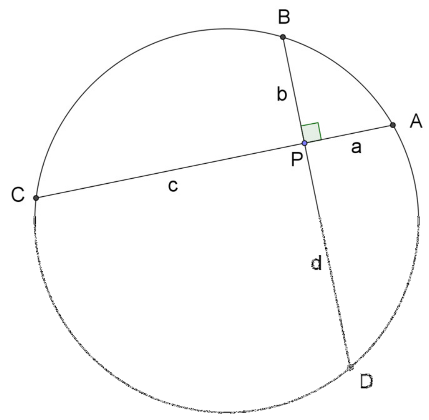


Fig. 7: Orthogonal chords

Plausibility consideration[3]

If two perpendicular blades of a "pizza knife" (intersection point *P*) are rotated (center *P*) by a small angle $\Delta\varphi$ then the area between the initial and the new position (grey in Fig. 8) is approximately

$$A \approx \frac{1}{2}\Delta\varphi\left(a^2+b^2+c^2+d^2\right) \qquad (1)$$

Explanation: We consider each of the four pieces as a real sector of a circle with angle $\Delta\varphi$ and the radii *a*, *b*, *c*, and *d*. The area of a sector with radius *r* and angle $\Delta\varphi$ is given by $b\cdot r/2=(r\cdot\Delta\varphi)\cdot r/2=r^2\cdot\Delta\varphi/2$.

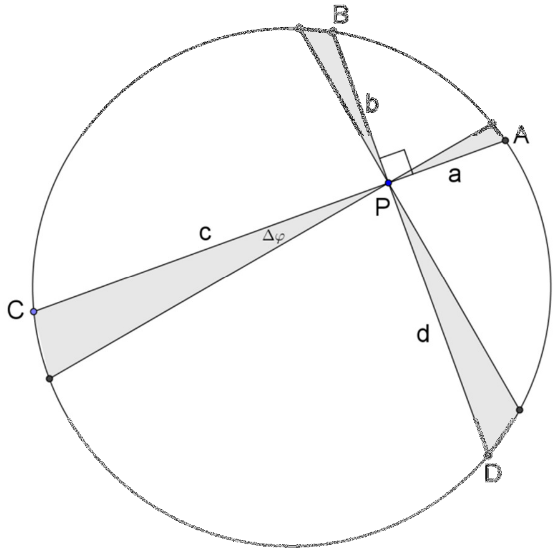


Fig. 8: Rotating further by $\Delta\varphi$

Here in this argumentation with plausibility we don't go into details why the approximate relation (1) also holds exactly (see "integral calculus" below), but one can see immediately: the real sectors with the radii *a* and *d* are a bit too big, the ones with the radii *b* and *c* a bit too small, therefore the approximation (1) will be rather accurate.

According to the above lemma $a^2+b^2+c^2+d^2=(2r)^2$ is constant and hence the *area* of the grey parts (Fig. 8) is *proportional* to the *angle of rotation* because the factor $\Delta\varphi/2$ is independent of the position of *P* and of the initial position of the two perpendicular blades. When we think of this rotation with $\Delta\varphi$ carried out many times we get the same proportionality, therefore there is no need to keep $\Delta\varphi$ small, it will work with any rotation angle φ. This proportionality can and should be affirmed with DGS. Such empirical results and findings with DGS of course cannot replace mathematical proofs but within arguments of plausibility they surely are a kind of affirmation that one is on the right way (if wanted one may look for a more rigid proof afterwards e.g. with integral calculus, see below). Hence the area of two perpendicular "double sectors" (as for instance the grey parts in Fig. 1 with $\varphi=\pi/4$) is $A=(1/2)\cdot\varphi\cdot(2r)^2=2r^2\cdot\varphi$, with $\varphi=\pi/4$ we get finally $A=r^2\pi/2$, that means that the grey parts together make exactly half the circle area.

Due to the mentioned *proportionality* (in some sense this is the mathematical core of this topic) we immediately get the following generalization: If the 90° quadrants are not only divided into two equal[4] parts of 45° – like above – but into say three 30° parts (in sum then we have 12 "sectors") then we may say with the same argumentation: With such a division one can divide a pizza equally between three persons, each person takes every third "sector". Again each person gets, in sum, two perpendicular pairs of opposite "sectors" ("double sectors"), the only difference: with an angle 30° instead of 45° (Fig. 9a: the first person gets the white double sectors, the second person gets the grey ones and the third person the black).

[3] I want to thank my friend and former colleague B. Schuppar (TU Dortmund, Germany) for useful hints.

[4] The "equality" is meant concerning angles.

The same procedure would work if we had n parts in every quadrant (in sum $4n$ parts), we would then have an equal division between n persons.

When we have an **even** number of "sectors" in each quadrant ($2k$, in sum then $8k$ "sectors") with this principle we have a possibility for an equal division of the pizza between $2k$ persons (each person gets 4 "sectors"). If we consider the $2k$ persons consecutively numbered (from 1 to $2k$) and if we think of the union of the "k even persons" and the "k odd persons" respectively then we again have an equal division into two subsets. In other words: If we number all the "sectors" from 1 to $8k$ (clockwise or counter clockwise) the area sum of the even ones is equal to the area sum of the odd ones!

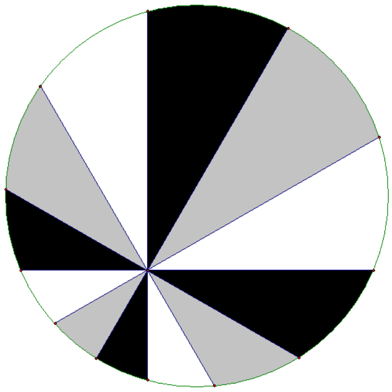


Fig. 9a: Equal division between three persons – three parts in every quadrant

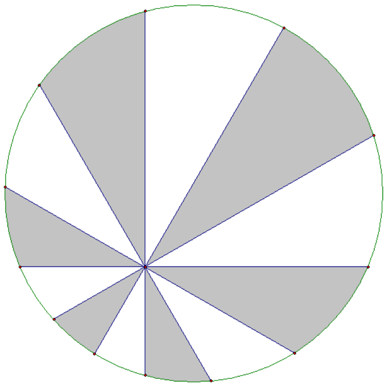


Fig. 9b: Equal division between two persons – three parts in every quadrant

But if every quadrant is divided into three parts (equal angles, see Fig. 9b; generally: an odd number of sectors in each quadrant) then our proof says nothing about the conjecture that still the area sum of the odd "sectors" is equal to the area sum of the even ones. In this case one does not have the easy situation of *pairs of perpendicular "double sectors" with equal color*. This case can be handled with calculus (see below). The phenomenon that the area sum is equal in both cases, adding all the odd "sectors" on the one hand and all the even ones on the other – regardless whether the quadrants are divided in an even or odd number of equal angles – , is called "pizza theorem" in the literature.

3 Calculus

Another possible treatment of this topic is given by an application of the **basic idea of integral calculus: Integrals are limits of product sums** $\sum_i f(x_i)\cdot\Delta x_i$. This useful and important basic concept for integrals – in many contexts – has been described quite often in the didactical literature (see e.g. Blum/Kirsch 1996, p. 62ff). With integrals we not only can calculate areas by using "slim stripes" (this geometric interpretation surely plays a big role but it should not be the only one) but they are useful in many other contexts. Often integral calculus at school is restricted to the application of the fundamental theorem of calculus $\int_a^b f(x)\,\mathrm{d}x = F(b)-F(a)$: Calculating integrals by using antiderivatives. Teachers and school text books often want to come to this result as quickly as possible in order to have plenty of possible calculation problems to solve for the students. The mentioned *theorem* is often even *degraded* to a *definition* like $\int_a^b f(x)\,\mathrm{d}x := F(b)-F(a)$ which in the sense of Hans Freudenthal can be seen as an "*antididactic inversion*". From the perspective of many teachers and school text book authors one may save several troubles and efforts (explaining what an integral *is*) but the prize to pay for that "advantage" seems clearly too high: students never get to know what an *integral* really *is*. A famous formula for the calculation of integrals – rightly named *fundamental theorem*! – is degraded to a *definition* in order to save time and possible troubles! If an explanation is more complicated just make a definition out of it? I think this must not be the kind of teaching that we should aim at!

In the following we also deal with calculating areas, not by slim rectangular stripes but by slim circular sectors, however the principle of product sums is the same. Of course the following cannot be expected as autonomous students' work but has to be explained by the teacher.

Different to Kroll/Jäger 2010 I propose to omit all the formal and technical aspects concerning calculating the area of the particular "sectors" because when adding the four areas one does not need any technical calculations by using the mentioned lemma of elementary geometry. By using integral calculus also the formal gap in the plausibility consideration in 2.3 is closed.

3.1 Reasoning with the Leibniz sector formula

Using the mentioned lemma of elementary geometry $a^2+b^2+c^2+d^2=(2r)^2$ and an application of integral calculus – also known as "Leibniz sector formula" – one can easily prove with calculus that the white and grey "sectors" in Fig. 1 have equal area sums (half the area of the circle).

Many different types of areas can approximately be seen as sums of slim circular sectors especially the above "sectors" that are very similar to real sectors. We are thinking of a fixed center *P* and a "radius ray" with variable length *r* (depending on the angle of rotation φ; "polar coordinates" but this explicit name is not necessary) and we are interested in the covered area of the "sector" that corresponds to the increase of φ by $\Delta\varphi$ (Fig. 10). The sector with radius |*PU*| is too small, the one with radius |*PV*| is too big for the area of the "sector" with the edges *P*, *U*, and *V* (this "sector" indicates the area increment when the corresponding angle increases from φ to $\varphi+\Delta\varphi$).

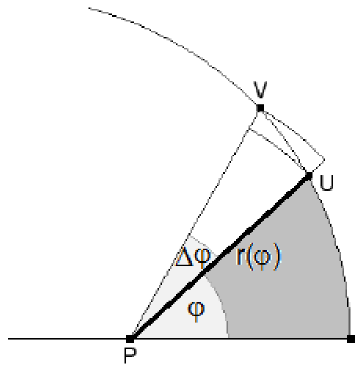


Fig. 10: Leibniz sector formula

For $\alpha \le \varphi \le \beta$ the covered area is divided in slim circular sectors. Adding up all these small sector areas yields a product sum which in the limit becomes an integral:

$$\frac{1}{2}\sum_i r^2(\varphi_i)\cdot\Delta\varphi_i \;\to\; \frac{1}{2}\int_\alpha^\beta r^2(\varphi)\,\mathrm{d}\varphi$$ where $[\alpha,\beta]$ is the domain of integration (concerning the associated angle φ, "Leibniz sector formula").

Applied to a white or grey "sector" in the pizza theorem we get $A=\frac{1}{2}\int_0^{\pi/4} r^2(\varphi)\,\mathrm{d}\varphi$ because the angle in each such "sector" is $\pi/4 = 45°$. We restrict to the grey "sectors" and draw thickly the "initial radii" $r_1(0), r_2(0), r_3(0), r_4(0)$ and radii in the end position – meant after the rotation by 45° – $r_1(\pi/4), r_2(\pi/4), r_3(\pi/4), r_4(\pi/4)$; we also draw a "position in between" of these radii (Fig. 11).

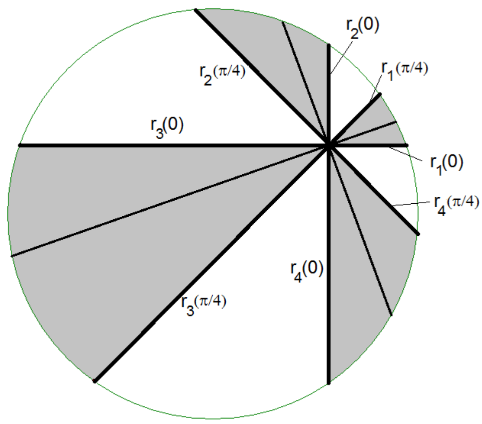


Fig. 11: Leibniz sector formula in the pizza theorem

For every angle $0 \le \varphi \le \pi/4$ we have: The corresponding radii are perpendicular to each other!

We need not calculate the particular areas of these "sectors" because of the above lemma from elementary geometry with r_1, r_2, r_3, r_4 instead of a,b,c,d : $r_1^2+r_2^2+r_3^2+r_4^2=(2r)^2$. With this we get:

$$\begin{aligned} A_{\text{grey}} &= A_1+A_2+A_3+A_4=\frac{1}{2}\int_0^{\pi/4} r_1^2(\varphi)\,\mathrm{d}\varphi+\frac{1}{2}\int_0^{\pi/4} r_2^2(\varphi)\,\mathrm{d}\varphi+\frac{1}{2}\int_0^{\pi/4} r_3^2(\varphi)\,\mathrm{d}\varphi+\frac{1}{2}\int_0^{\pi/4} r_4^2(\varphi)\,\mathrm{d}\varphi \\ &=\frac{1}{2}\int_0^{\pi/4}\underbrace{\left(r_1^2(\varphi)+r_2^2(\varphi)+r_3^2(\varphi)+r_4^2(\varphi)\right)}_{=(2r)^2}\mathrm{d}\varphi=\frac{1}{2}\cdot(2r)^2\cdot\frac{\pi}{4}=\frac{r^2\pi}{2} \end{aligned} \tag{2}$$

Remarks:

- Hereby the formal gap in 2.3 (plausibility consideration) is closed: Looking at (2) one can see immediately that the approximate formula (1) even holds exactly.

- Analogously here in the analytical perspective we easily see the mathematical core, the above mentioned proportionality: instead of $\pi/4$ this would work with every other angle φ of perpendicular "double sectors", one would get $A=\frac{1}{2}\cdot(2r)^2\cdot\varphi=2r^2\cdot\varphi$ as above in the plausibility considerations (using elementary geometry) in 2.3.

3.2 The boundaries of the pizza "sectors"

We have mentioned and proved already (with elementary geometry) that not only the pizza area but also the pizza boundary is divided equally by this method, this should be investigated now with calculus. The basic idea behind is the well known phenomenon that the circumference of a circle is the rate of change of the area with respect to the radius. For a proper understanding of the principle of the derivative it is very important to understand the mentioned phenomenon with regards to contents, not only formally by $\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\pi\right)=2r\pi$ (cf. Blum/Kirsch 1996, p. 61, Hefendehl-Hebeker 1998, p. 198f). Analogous considerations yield the insight that the surface area of a sphere is the rate of change of the sphere's volume with respect to the radius. These interesting issues can enrich the mathematics teaching of teacher students and students at school if not the syntax but semantics, contents, and understanding are in the focus of the teaching process.

We have already mentioned that – with arbitrary radii – the area of two "perpendicular double sectors" (as in Fig. 11 the four grey "sectors") is independent of the point's P position and of the position (concerning rotation) of the blades. The area only depends on the opening angle φ of the "sectors" (proportionality). Hence the same holds for the *difference* of two such areas, that is for the area sum $\Delta A=\Delta A_1+\Delta A_2+\Delta A_3+\Delta A_4$ of the four grey parts of the annulus when increasing the radius from r to $r+\Delta r$ (Fig. 12). Since

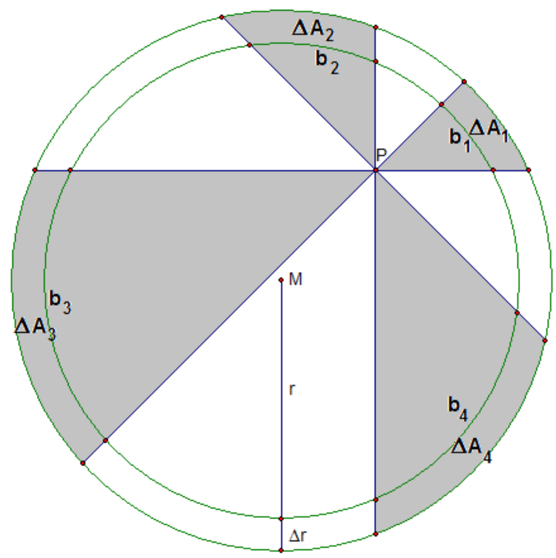


Fig. 12: Equal division of the pizza boundary

$\Delta A \approx (b_1 + b_2 + b_3 + b_4) \cdot \Delta r$ (the annulus everywhere has the same width Δr, hence these parts of the annulus together can approximately be thought of as a rectangle with length $b_1 + b_2 + b_3 + b_4$ and width Δr) does not change under translation or rotation of the cutter (see above) the same holds for $\frac{\Delta A}{\Delta r}$ and in the limit also for $\lim_{\Delta r \to 0} \frac{\Delta A}{\Delta r} = \frac{\mathrm{d}A}{\mathrm{d}r} = b_1 + b_2 + b_3 + b_4$ (sum of the four arc lengths of the grey "perpendicular double sectors"). That means the sum of the four arc lengths $b_1 + b_2 + b_3 + b_4$ on the pizza boundary is independent of the point's *P* position and of the rotation position of the blades. This sum itself therefore is proportional to the opening angle φ. By this we have shown in an elementary way that for an opening angle of $\varphi = 45°$ we get exactly half of the circumference of the circle – an equal division of the pizza boundary between two persons (analogous if we have *n* persons and an opening angle $\varphi = 90° / n$ of the "sectors").

3.3 The pizza theorem when each quadrant is divided in an arbitrary number of "sectors" (odd or even)

With the considerations so far we have shown in two ways (using elementary geometry on the one hand and calculus on the other) that the pizza theorem holds in the case of dividing each quadrant in an even number of pieces ("sectors"): The area sum of the grey "sectors" equals the area sum of the white ones (in each case we have half the area of the circle). In the following we will show that this is the case in general (also when we have an odd number of pieces in every quadrant).

One could ask here: Why did we not use this method (general case) from the very beginning? The answer is easy: The mathematical core of the above reasoning (the mentioned proportionality) is very important for properly understanding the phenomenon, but this proportionality does not appear in the following considerations. A second reason could be that in the one or another situation the general case may be not so interesting, that one wants to reduce complexity and deal only with our initial special case of four cutting blades. For this case we wanted to provide possibilities.

We have to prove that the sum of all the areas of the odd "sectors" is exactly half of the circle area. We will have to calculate sector areas by corresponding integrals (in the above version we actually did not do real calculations). We will not determine areas of single "sectors" but of two opposite ones, so called "double sectors" (see Kroll/Jäger 2010).

First we have to deal with two important items that we will need:

- $$\sum_{k=0}^{m-1} \sin\left((2k+1) \cdot \frac{\pi}{m} \right) = 0 \qquad \text{and} \qquad \sum_{k=0}^{m-1} \sin\left(2k \cdot \frac{\pi}{m} \right) = 0 \qquad (3)$$

  Proof: If we write $\frac{2\pi}{2m}$ instead of $\frac{\pi}{m}$ these trigonometric equations become quite clear: $\frac{2\pi}{2m}$ is the central angle of a regular (2*m*)-gon (*even* number of vertices!), e.g. drawn in the unit circle with one vertex at $(1|0)$ and the opposite one at $(-1|0)$ (Fig. 13, regular octagon). $(2k+1) \cdot \frac{\pi}{m}$ are the odd multiples of this central angle and the corresponding vertices of the regular (2*m*)-gon lie symmetrically with respect to the x-axis. Thus in sum the corresponding sin values cancel each other (analogous in the case of $2k\cdot\frac{\pi}{m}$, the even multiples of the central angle).

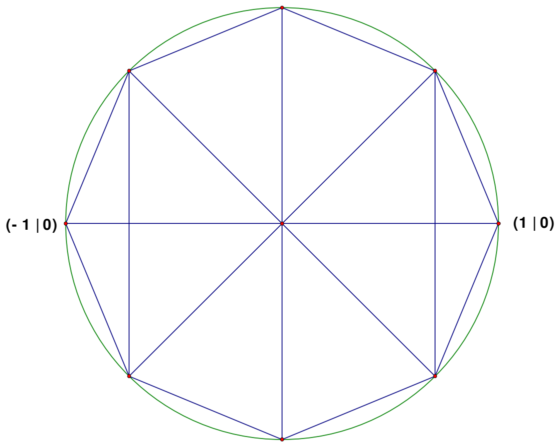


Fig. 13: Regular polygon with an even number of vertices

- Let $r_1(\varphi):=|PS|$ and $r_1'(\varphi):=|PR|$ be the distances from P to the perimeter of the circle (Fig. 14). One can read off immediately $f:=|QP|=e\cdot\cos\varphi$ and $h:=|MQ|=e\cdot\sin\varphi$. Because of symmetry reasons we have $c:=|QS|=|QR|$ and with Pythagoras we get $c^2=r^2-h^2=r^2-e^2\cdot\sin^2(\varphi)$. Due to $r_1'=c+f$ and $r_1=c-f$ we can write

$$
\begin{aligned}
(r_1')^2+r_1^2 &= 2\left(c^2+f^2\right)\\
&= 2\left(r^2-e^2\cdot\sin^2(\varphi)+e^2\cdot\cos^2(\varphi)\right)
\end{aligned}
$$

and finally: $(r_1')^2+r_1^2=2\left(r^2+e^2\cdot\cos(2\varphi)\right)$

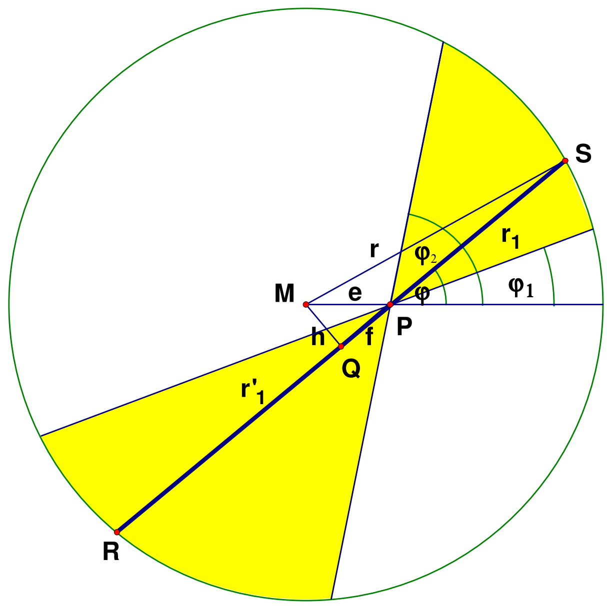


Fig. 14: Leibniz sector formula in a "double sector"

With this knowledge it is not difficult to calculate the area of a double sector by using the Leibniz sector formula. We get (see Fig. 14):

$$
\frac{1}{2}\int_{\varphi_1}^{\varphi_2}\left[r_1^2+(r_1')^2\right]\mathrm{d}\varphi=\int_{\varphi_1}^{\varphi_2}\left[r^2+e^2\cdot\cos(2\varphi)\right]\mathrm{d}\varphi =
$$

$$
r^2(\varphi_2-\varphi_1)+\frac{1}{2}e^2\left(\sin(2\varphi_2)-\sin(2\varphi_1)\right) \qquad (4)
$$

We apply this formula to the particular double sectors, e.g. in Fig. 15 we have altogether n = 6 such double sectors (the odd ones are white). This number n is even at any rate (n = 2m) independent of the fact whether each quadrant is divided into an even or odd number of "sectors".

In the i-th "double sector" the "polar angle" ranges from $\varphi_1=(i-1)\cdot\frac{\pi}{\underbrace{n}_{2m}}$ to $\varphi_2=i\cdot\frac{\pi}{\underbrace{n}_{2m}}$.

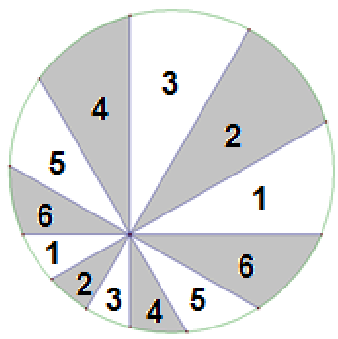


Fig. 15: Pizza theorem with three "sectors" in each quadrant

Thus from (4) we get for $|DS_i|$, the area of the i-th "double sector":

$$
|DS_i|=r^2\cdot\frac{\pi}{n}+\frac{1}{2}e^2\left(\sin\left((2i)\cdot\frac{\pi}{\underbrace{n}_{2m}}\right)-\sin\left((2i-2)\cdot\frac{\pi}{\underbrace{n}_{2m}}\right)\right)
$$

$$
=r^2\cdot\frac{\pi}{n}+\frac{1}{2}e^2\left(\sin\left(i\cdot\frac{\pi}{m}\right)-\sin\left((i-1)\cdot\frac{\pi}{m}\right)\right)
$$

Now we calculate the area sum $\sum_{k=0}^{m-1}\left|DS_{2k+1}\right|$ of the odd “sectors” with the numbers $i = 2k+1$ and do the summation of the particular parts separately: We have $m = n/2$ times the summand $r^2 \cdot \frac{\pi}{n}$ which yields $r^2 \cdot \frac{\pi}{2}$, that is exactly the half of the circle area. The other trigonometric sums yield 0 according to (3), and by this the general pizza theorem is proved.

Finally another pizza theorem (http://mathworld.wolfram.com/PizzaTheorem.html), easier to prove: The volume of a pizza with radius z and thickness a is given by Pi z z a .

4 Teaching aspects

All in all this is an ambitious topic with many possible connections to other mathematical fields. It seems to be a good opportunity to foster elementary geometry in mathematics teaching by dealing with an interesting phenomenon (equal division of a pizza). In German speaking countries in my opinion there should be payed more attention to elementary geometry in lower secondary schools. Here one may think of matters like symmetry, reflections, rotations, division in pairwise congruent pieces (equidecomposability); these learning matters can be strengthened, extended, and linked when dealing with this topic. In the teaching of differential and integral calculus in school (high school, college) often syntax and special techniques of calculation (things that often could be done by computers) are given priority while semantic aspects and real understanding are neglected. But within the presented topic the idea of an integral as a limit of a sum of products (areas as sums of slim circular sectors) is constitutive, and this is a really semantic aspect. New terms like “polar coordinates” or “Leibniz sector formula” etc. are not important. So far to competencies with regards to contents. Competencies with regards to processes that can be fostered by the suggested topic are: cross-linking matters, problem solving, exploring situations, verifying phenomena (using Dynamic Geometry Software – DGS), argueing and reasoning.

How and in which grades could this phenomenon be dealt with at school?

In lower grades (say grade 7 to grade 9) a similar problem with a square instead of a circle may be dealt with as a problem that students are supposed to solve individually. Maybe some smaller hints from the teacher are necessary?

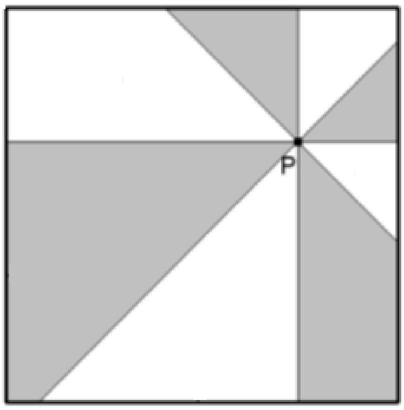


Fig. 16: Square and cutter

In a square the knife (cutter) is placed like in Fig. 16: two blades parallel to the edges and the other two parallel to the diagonals. Find arguments to show that the area sums of the white and grey pieces respectively are equal (cf. Kroll/Jäger 2010, 103f).

Remark: If P is not the center of the square (in the case of the center everything is clear) then P lies in a special quadrant. Without loss of generality we may assume that P lies in the right quadrant above.

Here there are several possibilities to see that the area sums are equal. When having in mind individual work of students it is always a big advantage if there are several ways to solve a problem because in this case different approaches can lead to the solution. A possible hint for a short and elegant solution could be: "construct a special line segment (dashed) – the upper edge reflected on the horizontal line segment through *P* (Fig. 17)".

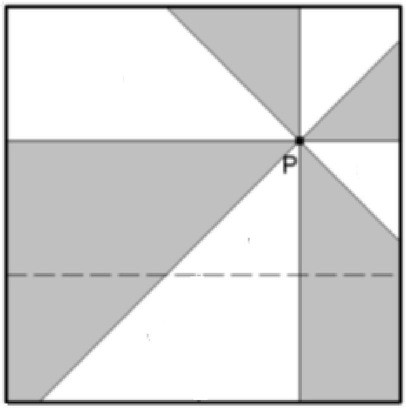


Fig. 17: Square and cutter – first hint

With this hint the following is clear by symmetry: above the dashed line the white parts equal the grey parts, therefore one has to think only of the parts below this line. In the case that students cannot find a solution individually despite of this hint there can be another one: Draw another line segment which leads to equal areas III and IV. Then the problem is reduced to finding reasons why I and II have equal areas (Fig. 18). If lower grade students succeed in reasoning this fact it can be seen as a good performance (here 45° angles and the resulting symmetries can play an important role).

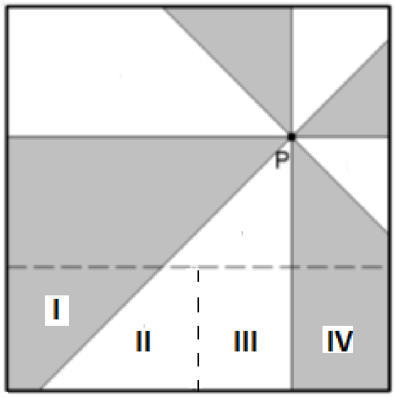


Fig. 18: Square and cutter – second hint

Concerning the original problem (circular pizza) in secondary schools students could gather practical experiences on the one hand (cutting circles of carton in the described way and weighing all the grey and white pieces[5]) and on the other the phenomenon could be explored by using DGS as a means of measuring the corresponding areas.

Also dealing with the dissection following P. Gallin (Fig. 5) seems to fit well to secondary school (grade 7 to 9), maybe even as problem to be solved autonomously by groups of students. Fig. 5 probably has the character of a "proof without words" even for many students of grade 7 – 9. I do think it could work with these students because one does not need more ingenious ideas, the helpful octagon is drawn in the figure, it is also explained how it arose. Also not so capable students can have the idea to simply count the area pieces $a, \ldots, g, x$ (their congruency is here immediately clear) in the white and in the grey zone outside the octagon. Also inside the octagon the way of reasoning is evident. The proof for the equality of the pizza boundaries following P. Gallin (using elementary geometry, see Fig. 6a,b) is also suitable for secondary school students. If the problem reduction Fig. 6a → Fig. 6b is done by the teacher (students in secondary schools will probably not see this possibility by themselves) then the further handling of the problem could be done autonomously in groups (maybe small hints from the teacher are necessary). For better supporting ideas and concepts and for visualizing the use of DGS (as a means of measuring) could be recommended, best in advance of the proof.

In high school or college (when the basics of calculus are taught) the phenomenon could be presented by the teacher, independent students' activities seem not to fit here. When realizing this ambitious program the main goal would be the important idea of the integral as a limit of product sums (here: slim circular sectors). In school the aspects of 3.1 would suffice, the ones of 3.2 (boundaries of the pizza pieces) and 3.3 (general case) are not necessary there. The important lemma of elementary

[5] If the grey pieces together have the same weight as the white pieces together then it is clear that the area sums are equal, too.

geometry can provide aspects of crosslinking several mathematical fields (calculus, elementary geometry), also individual students' work is well possible (problem solving, maybe in groups).

Also at university – teacher education – Fig. 5 (dissection following P. Gallin) could be dealt with, students of a geometry course could analyze the figure autonomously. But at this level students could also analyze Fig. 4 (including the helpful octagon; Fig. 3 – without this octagon – is with regards to my experiences not so good for that purpose). For preparing such an analysis of Fig. 4 one could pose the following problem (then working with the octagon is easier):

A rectangle *ABCD* is rotated (center *M*) by 90° (→ $A'B'C'D'$) so that the octagon $A'BCB'C'DAD'$ is established (see Fig. 19). Explain exactly why e.g. the line segments $A'B, D'C, B'C$ are "45° lines" (for reasons of clearer arrangement only one diagonal $D'C$ is shown, but the others are such "45° lines" as well).

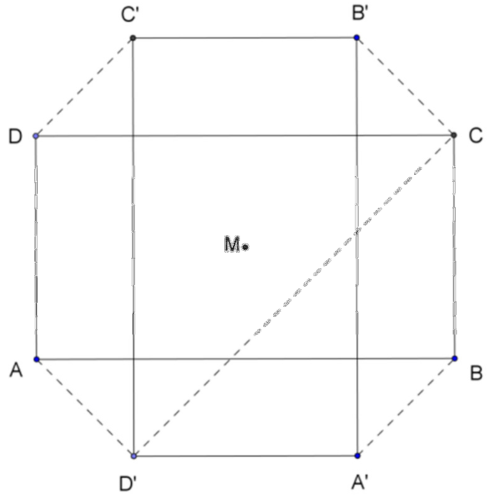


Fig. 19: 90° rotation of a rectangle

It is clear that in calculus courses at university (teacher education) all here presented aspects can be covered. But I have made the experience that the Leibniz sector formula is very rarely presented is such courses. I regret that, not because I want students to solve many integrals in polar coordinates, I regret it primarily because I think having another geometric idea of product sums (not only slim stripes and rectangles, but also slim circular sectors) is a valuable enrichment of the idea of integrals in mathematics teacher education. As I got to know it for the first time I asked myself: Why did nobody during my studying time at university (in calculus courses) tell us about this simple and important idea? Dealing with the general case (cf. 3.3) is in my opinion even at university not obligatory.

References

Blum, W., & Kirsch, A. (1996). Die beiden Hauptsätze der Differential- und Integralrechnung. *mathematiklehren*, *78*, 60–65.

Carter, L., & Wagon, S. (1994). Proof without Words: Fair Allocation of a Pizza. *Mathematics Magazine, 67*(*4*), 267. For a new colored version see Wikipedia: http://en.wikipedia.org/wiki/Pizza_theorem

Gallin, P. (2011). Exzentrische Kuchenhalbierung. *Bulletin des VSMP* (*Verein Schweizerischer Mathematik- und Physiklehrkräfte*), *116*, June 2011, 11–19. Online: http://www.gallin.ch/KuchenhalbierungBulletin.pdf

Hefendehl-Hebeker, L. (1998). Aspekte eines didaktisch sensiblen Mathematikverständnisses. *Mathematische Semesterberichte*, *45*, 189–206.

Herget, W., Jahnke, T., & Kroll, W. (2011). *Produktive Aufgaben für den Mathematikunterricht in der Sekundarstufe II* (pp. 20 and 100–105). Berlin: Cornelsen.

Kohnhauser, J.D.E., Velleman, D., & Wagon, S. (1996). *Which Way Did the Bicycle Go?* Dolciani Mathematical Expositions 18, Washington: Mathematical Association of America.

Kroll, W. & Jäger, J. (2010). Das Pizzatheorem. Ein Thema mit Variationen. *mathematica didactica*, *33*, 79–112. Online: http://www.mathdid.ph-gmuend.de/documents/md_2010/md_2010_Kroll_Jaeger_Pizzatheorem.pdf

Nelsen, R. B. (2004). Proof without words: Four Squares with Constant Area. *Mathematics Magazine*, *77*(*2*), 135.

A Conversation with Herbert Tate: Mathematics Educator and Builder

Christian Genest (@mcgill.ca)
McGill University, Montréal (Québec) Canada

Abstract: *Herbert Tate was a Professor of Mathematics at McGill University (Montréal, Canada) from 1921 to 1964. As the author of four textbooks, and in his capacity as Chairman of the Department of Mathematics from 1948 to 1960, he played a key role in structuring the institution's research and study programs in mathematics during an important period of growth. McGill's current position as a hub of mathematical research owes much to him. In this interview given shortly after his retirement, Herbert Tate describes his career and shares some of his views about mathematics and related topics. Beyond its archival value, this interview reminds us of the extent to which infrastructures and mentalities have changed, at least in Canada, over the past century.*

Keywords: Mathematical education · Mathematical history · McGill University

Mathematics Subject Classification: Primary 01A72; Secondary 01A70

1. Introduction

Herbert Tate studied mathematics at Trinity College, Dublin. After completing a Master of Arts, he moved to Montréal, Canada, in 1921 to teach mathematics in the newly created School of Commerce at McGill University. A pioneer of actuarial education on campus, he wrote four textbooks summarizing some of his teachings: *Interest, Annuities and Bonds* (Tate, 1929), *Elements of the Mathematical Theory of Interest* (Tate, 1937), *Elementary Mathematical Analysis* (Tate, 1946), and *Mathematical Theory of Interest* (Tate, 1947).

Herbert Tate was a scholar but not a researcher by today's standards. His only entries on *MathSciNet* are solutions to problems in *The American Mathematical Monthly* published between 1936 and 1938; see Gelbart & Tate (1936), Claudian & Tate (1937), Levenson, Locke & Tate (1938), Musselman, Tate & Sparrow (1938), and Underwood & Tate (1938). Nonetheless, Herbert Tate was highly influential, both as an educator and as a builder of mathematics at McGill. He chaired the Mathematics Department from 1948 to 1960, in a period of rapid growth. Through his work and vision, the department got a head start in research on the Canadian scene and developed an enviable worldwide reputation that it holds to this day. Professor Tate retired in 1964 and was made Emeritus the following year. McGill still runs a loan fund in his name that gives special consideration to “those who, like the donor, begin a university education as mature students and require financial assistance.”

Below are excerpts from an unpublished interview of Herbert Tate that Lorne Gales and Albert Tunis conducted at McGill on June 2, 1964. It is one of a handful of interviews with academic staff available only in reel-to-reel audiotape format from the McGill University Archives Audio Collection. The original quotes are verbatim but the material was reorganized somewhat to improve the flow. I also added occasional words and a few details in square brackets for clarification. Information about people mentioned in the interview is also provided wherever possible either via footnotes or Web links accessible by clicking on names in blue.

Beyond its archival value, this interview reminds us of the extent to which both the infrastructures and the mentalities have changed, at least in Canada, over the past century. This paper is intended as a tribute to a man of his time, modest and efficient, who truly believed in mathematics education and devoted his entire life to it, with much success and — to this day — little recognition.

2. Recruitment

Q: Professor Tate, I’d just like to take you back a bit in years. I find that rather interesting in talking to members of Faculty who’ve been here for some time to find out how they first came to come here. You know what I mean.

T: Well, that’s very interesting. I was brought to McGill primarily through the efforts of Professor R.M. Sugars [Associate Professor of Spanish], who was Director of the School of Commerce. He wanted somebody to take over mathematics in the School of Commerce and at that time — it was after the First World War — there was a world shortage of mathematicians. And he went to the Chairman of the [Mathematics] Department, who in those days was Professor James Harkness, and asked him if the department could get somebody to take charge of mathematics in the School of Commerce. I understand that Professor Harkness said it was impossible; he couldn’t even get some people for his own division. Well then, Professor Sugars said: “If that is the case, if you leave it in my hands, I will guarantee to get somebody.” So he wrote to his old university, Trinity College in Dublin, and to the Vice-Provost [Thomas Thompson Gray], who, by the way, was a very famous classical scholar, who sent it on to the Professor of Mathematics who had been my tutor, and he said: “If you wish to apply for this post, I will not approach anybody else.” And that is the way I came to McGill.

I was appointed Assistant Professor in 1921. Now, in those days, the future of the School of Commerce was uncertain, as it is today [1964]. It always has been uncertain. Nobody has ever known whether it should be in the university or not, or whether, as it is here, whether we should keep on with it. And the letter I got from the Principal, Sir Arthur Currie, mentioned that my immediate and probably permanent work would be with the School of Commerce. And when I came out, I asked him what was the meaning of that. “Well,” he said, “we’re not quite sure whether the School of Commerce would be here in another two or three years. But,” he said, “you’ll be all right! You’ll be all right for that reason,” he said, “still we have made you a member of the Mathematical Department, rather than a member of, solely, the School of Commerce.”

3. Development of Mathematics at McGill

Q: Professor Tate, from your vantage point now, after 43 years, I wonder if you could look back and indicate to us, broadly, how mathematics at McGill has developed from when you first came.

T: Oh yes. When I first came to McGill in 1921, the Mathematical Department was quite small. It was, for all practical purposes, in two divisions. Arts and Science with a Full Professor at a head, Professor Harkness, who was a very distinguished Cambridge mathematician, and Engineering, with a Professor of Applied Mathematics, Professor Daniel Murray, at a head of that. Now, the Department as a whole only met together when matters of common interest were at stake. But for all practical purposes, it was two separate departments.

When Dr. Harkness died, I thought 1923, 1924 [December, 1923], Sir Arthur Currie, who, after all, remember, was a very distinguished military man and liked centralization, said: “There is not two different kinds of mathematic as I see it, there is only one.” So he made one department under the chairmanship of Dr. Murray. And then Dr. [Charles Thomas] Sullivan — a very distinguished geometer, as you know, who was [here] so many years [1908–48], and did his best to make mathematics as easy as possible for unfortunate students — came over, transferred from Engineering to the Faculty of Arts, and became Peter Redpath Professor of Mathematics. When Dr. Murray retired, in about 1928 or ‘29, Dr. Sullivan succeeded him. And he was Chairman, oh, until about 17 years ago[1].

Q: I was gonna say, after the war, because he was here and we ran a picture of him in the *McGill News* at the time he retired.

T: Now, in the early days, when I joined in 1921, the normal load of teaching in mathematics was 16 hours a week for members under the rank of Full Professor. Full Professors did 12. We had no offices. There was a general room in the Faculty of Arts and Science in the old Arts Building, with pigeonholes for our papers. It served as a Faculty room, it served as a combination room, and it also served as the Dean’s office. The only person, in those days, speaking roughly, who had an office, was the Dean’s secretary. I don’t even know that the Dean had an office. I used to see him sitting at the head of the table in the Faculty room, and when he wanted a letter, he went into his secretary’s office and dictated it there.

Well, that made it very difficult to carry on anything in the nature of research work, or sustained study, because in those days, the timetable was made up by the Dean’s secretary, and all she was interested in was finding a place somewhere, and seeing a vacant space, and putting somebody in to lecture there. Consequently, we might have a lecture at 9 o’clock in the morning; the next might be at 1 o’clock in the afternoon, and the next at 5 o’clock in the evening. We had nowhere to go, we had nowhere to sit. The library was overcrowded and, as I say, you would go into the Faculty room, expecting to pick up your papers, you would find a Faculty meeting in progress, and you had to get out as quickly as possible.

Now, one of the greatest advantages that the staff had when the new Arts Building was made was that we had offices. And we were no longer lost sheep wandering from the library to the Faculty room, down to Murray’s, spending too much time there drinking a cup of coffee, and in those ways, our whole time was frittered away. Then, due to the efforts of Professor [Stephen] Leacock, a new system of timetabling was introduced. Before this was introduced, some courses had four hours, some courses had two hours, and some courses had three hours. There was no uniformity. He introduced a uniform system of working in three-hour units. If your first lecture was at 9 on Monday, your next lecture was at 9 on

[1] He was at McGill from 1908 to 1946, but was replaced as Chair in 1945 by A.H.S. Gillson, who became President of the University of Manitoba in 1948.

Wednesday, and your next lecture was at 9 on Friday. And he firmly believed, and I think he persuaded all of us, that if you could not teach a course in three hours you either could not teach it at all or it was not worth teaching.

Q: I cannot think of anything more amusing, Professor Tate, than Professor Leacock with Miss Field. Because obviously, she must have been in charge of the timetable that you refer to, and I remember Miss Field as a tyrant in the Arts Building.

T: Yes, yes. Well, over the years, the next thing was, our hours at that time were reduced from 16 to 12 hours a week, and this reduction has gone on until today. I suppose, in the Mathematics Department, the average lecturing hours is about 8 hours a week, giving people time for reflection, proper preparation of lectures, for study and research.

4. Curriculum

Q: How about the actual curriculum itself? I suppose that expanded tremendously over the years.

T: The curriculum has expanded tremendously. When I came over in 1921, and frankly until about 18 years ago [1945–46], the undergraduate curriculum in mathematics consisted of 13 courses, almost all of them Honors courses. There was a big general course for first-year students, Mathematics 1. And there was a course for whom we considered the brighter students and who would eventually take Honors Mathematics 1A, 1 Advanced. Mathematics 1 consisted of all first-year students, roughly speaking, except about 15 or 16 who were in the first-year advanced course.

Now, after that, the advanced course, their future was clear. If they went on to Honors in Mathematics, they took courses 3 — famous calculus course — 4, 5, 6, up to Mathematics 11. But for the general students there was really nothing. There was a Mathematics 2, which really was a remedial course, which had to be offered in order to make teachers conform to the regulations of the Department of Education, and it was really meant to appeal to first-year students who were going in for education but who did not attain 65 % in their first year. Other general students who wished to take courses had to take the courses offered to the Honors students. Not many of them took it and very few of them survived. And it was not until the demand grew up from other departments to have courses tailored to their requirements not on an Honors basis... For example, 25, 20 odd years ago, I introduced Elementary Calculus. That was when Professor Gillson was Chairman of the department. And he said: “Well, if we are willing to have this, if you could guarantee we would have 18 students or so.” Well, we started off with 35 students; today there are 500 in that course.

Now, today, we offer over 30 courses in the undergraduate curriculum. We offer courses on the ordinary mathematics: analysis, calculus, advanced calculus. We also offer series of courses designed to meet the requirements of Honors students in chemistry; there is one course every year: first, second, third, and fourth year. The Department of Economics requires courses in statistics. We have courses in statistics, we have courses in actuarial mathematics, we have courses in group theory, which 20 years ago were regarded as being specifically for graduate students. In our graduate work, until about 18 or 19 years ago, we offered about six or seven courses in the Graduate School, most of which were never given. I would say that for about 20 years, up from 1921 up to about 1952, the number of graduate students we turned out in mathematics was only five or six… in 20 years.

Q: How does that compare with today? [1964]

T: Today, we have 40 students in the Graduate School alone each year, half of them Masters and half of them Doctors. Until Professor Gillson — and I succeeded him — took over, we had no Doctors candidates at all in mathematics. They all had to go elsewhere.

5. Creation of the Graduate School

[Q: Until 1945, Mathematics was almost wholly a service department with only seven faculty members. Can you say a few words about the creation of the Graduate School in Mathematics?]

T. When we decided to set up a Graduate School in Mathematics, I discussed the matter with Professor [Horace Noel] Fieldhouse, [a historian] who was Dean of the Faculty of Arts, and, of course, Doctor [Frank] James, who was Principal, and Doctor James told me: “Well,” he asked me, “have you got a plan?” Well, I said: “I have a plan but it is not fully matured at the moment.” “Well,” he said, “come back to me when it is. We will discuss the things.” So I discussed the matter with Dean Fieldhouse and Doctor James, both of whom gave us every possible encouragement and help. And we decided that the best thing to build the Graduate School is from the top and not from the bottom. And we decided we would go out and get three or four really first-class mathematicians of a senior rank.

The first department hiring we made [in 1948] was Professor [Hans] Zassenhaus, who was a very world-famous algebraist and group theorist and whose work will probably be the foundation on these subjects for the next hundred years. He came in and developed our Graduate School and helped us enormously. And then Professor [Wacław] Kozakiewicz, a Pole from the University of Warsaw[2], who put our Statistical Section on a firm basis and who also was a very versatile analyst. And Professor [Charles] Fox, who is a remarkably versatile man who has written on practically every face of mathematics. And, of course, we have our present Chairman, Professor [Edward] Rosenthall, who was with us at that time and also is an expert in number theory. Without the help of these people, in fact, without these people, our Graduate School would never have even gotten off the ground. But from that day, it has never looked back and it is probably very, very well known today. Now, some of the... Oh, and I should mention, of course, in Theoretical Physics Professor Phil Wallace, whose section, really, was one of the brightest spots in the Mathematics Department.

6. Successful Students and Other Graduates

Q: Do you remember many of your students as outstanding, as brilliant students? Have you seen how they progressed in life?

T: Yes, all my students are outstanding, in one respect or another. In fact, there are so many of them that it is invidious to mention by name. There is [the interviewer] Mr. Lorne Gales, for example [chuckle]. Graham Glasgow, whom you all know, was a member of perhaps the most brilliant class in commerce I have ever had. In the second-year commerce class, there were 90 students. And my difficulty that year was not to pass them, but to keep most of them out of the first class. He was one of them. There are many others as well.

Anston McKim was another student, and Ken Carter, who is... Isn't he head of the Taxation Committee?[3] And there were, at one time, four people who were the Senior Auditors of the provinces of

[2] He was previously affiliated with the Université de Montréal.

[3] The Royal Commission on Taxation, or Carter Commission, was appointed in 1962 by the Prime Minister of Canada, John Diefenbaker, to examine and to recommend improvements to the entire federal taxation system; it produced an influential 6-volume report in 1966.

Alberta, Ontario, and so on. Now, there is Walter Markham, who is President, and there is David Scott who is the Vice-President of, the Continental Life Insurance Company. And there is Arthur Weaver, who is Vice-President of the John Hancock Insurance Company in Boston. There is Vernon Lawson, I think, who is... You can check this. I think he is Vice-President of the Montreal Life. And there is Anderson — I have forgotten his Christian name — who also is there.

Q: You taught a lot of actuaries! [Paradoxically, there is now no actuarial program at McGill.]

T: I taught a great many actuaries. That has been my specialization. And another man is Roy Saunders, who is controller of the Imperial Life and may be even manager now.

Now, some of our graduates: these are not all members of our Graduate School, but they all took either primary or subsequent degrees in McGill. I mention Professor Louis Nirenberg, who is in the Mathematical Institute of New York and who is one of the most distinguished men living in differential equations. He received the Bôcher [Memorial] Prize two years ago [in 1959, actually] in mathematics from the American Mathematical Society, which is the blue ribbon of mathematics all over the world. I [also] mention Professor Raoul Bott, who started with us in Engineering, then went to the Carnegie Institute, where he graduated with the PhD in theoretical physics, and is now Professor of Pure Mathematics in Harvard at an extraordinary early age and is one of the five or six really outstanding topologists in the world.

Professor [Viktor] Linis, who is Head of the Department of Mathematics in the University of Ottawa; Professor Norman Smith, who is Head of the [Mathematics] Department at Sir George Williams College [now Concordia University, Montréal]; Professor Raymond Ayoub, who is Professor of Mathematics at Pennsylvania State University and who has just finished his commission for writing the definitive book on the analytical theory of numbers on behalf of the American Mathematical Society; Professor Norman Oler, who has just been appointed to the staff of the University of Pennsylvania; Professor George Cree in Alberta, and Professor [Donald] Betts, who is in Theoretical Physics in Alberta; and Professor [John David] Jackson, who was on our staff from the University of MIT for six or seven years and is now Professor of Physics at the University of Illinois, who is an extraordinary, able theoretical physicist and with Professor Wallace gave us tremendous appetence in our Department of Theoretical Physics.

7. On Women in Mathematics

Q: A number of years ago, Professor Tate, you sat in at a meeting of a committee that was studying the possible revision of the curriculum of the School of Commerce, and I was particularly struck by a statement that you made; two statements really. The first was a statement on what you thought of the importance of extra-curricular activities among students. The other statement, which I would like you to repeat again today, are your reflections on the difference between men and women as they advance in the study of mathematics. I wonder if you recall that latter statement.

T: I am afraid I don't.

Q: It was something to the effect that after the first couple of years, men and women, from a mathematics point of view, are just about the same, but after that — when came a little more of a philosophy of mathematics, I think it was — that the woman's mind seemed to differ. Admittedly, this is almost... this is 12–15 years ago you said that [i.e., around 1950].

T: Now, do you want to get me into trouble with my wife?

Q: No, this is just for amongst us boys at the moment. We cut the tape.

T: Yes...

Q: And on your thoughts on teaching them at McGill because you have taught an awful flock of them.

T: Well, women work harder than men. And they are more conscientious than men. But, very often, a woman's thoughts are on something else. And there are — we have had — very, very good women in mathematics; we have had quite brilliant women in mathematics. But I think, speaking by and large, they have never reached the top heights, which men reach. Because of course they haven't got the spar of ambition. Now, many of our women have gone out to be married. They have done extremely well, they have got Master's degrees, Doctor degrees, and then they go out to be married and that is the end of it.

As one former woman student of mine said, who was a very brilliant mathematician indeed in her early days in McGill — she was first in every year, and I met her a few years ago, she is married and she has had a couple of children; she married an engineer as a matter of fact — and she said: "When I married my husband, I knew far more mathematics than he did. After all," she said, "he was only an engineer!" But she said: "Today, I can hardly remember enough to make up my accounts! The reason is," she said, "I, in a busy life, I just simply have lost all my techniques and I haven't had time to keep them up."

Now, when a woman pursues her profession single-handed and single-mindedly, very often she does as well as any man and she has reached the first rank. There is one very well noted English woman mathematician, Miss [Mary] Cartwright, who is Mistress of — Head of — Girton College in Cambridge, an extremely well-known mathematician, and, of course, there have been others: Emmy Noether, a famous German mathematician, and so on. But as I say, after all, one of the things that they don't, and by and large, pursue is their mathematics once they have married. And, after all, that is the ultimate goal for most women.

8. Student Training and Mathematical Talent

Q: Professor Tate, there is one other interesting point. In 1961, when McGill submitted a brief to the Royal Commission on Education, one of the statements that was made in one section of the brief was the fact that the university was somewhat concerned over the fact that the young students coming along out of the high schools seemed to evade the more challenging subjects, and I think mathematics and physics were named. Have you any comment on that?

T: Well, it is very difficult to see how they can evade them because in effect they are compulsory in the first year!

Q: I think probably the point was being made about specializing or Honoring, or taking these on as their special course.

T: Yes, well, mathematical ability is quite rare. I mean outstanding mathematical ability is quite rare. Outstanding ability to write poetry is rare, outstanding musical ability is rare. And for that reason, the number of students who are really capable of doing advanced work in mathematics, at any time, is a very small proportion of the whole.

Q: I think what I'm getting at, really, is I wanted to get some comment from you about whether you thought that the students coming in were adequately prepared in these things on the lower level so that it gave you something to work with.

T: Well, we would all like students to be better prepared and we all are inclined to blame the other man for their defects. But I think that the students in Québec are as well prepared as they are anywhere on this continent. As a matter of fact, I had some experience of examining on the other side [in Europe] at Matriculation Level and I have been examining for the School Leaving Certificate in McGill — at least in McGill and in the Department of Education — for about forty years, and I must say that I would regard the average as good as anywhere in the British Empire, and probably better than most places in the [United] States.

Q: Do you think that one is born with mathematical ability or is it teaching?

T: Oh, undoubtedly. It is innate. It can be developed by teaching but if it is not there, you can't develop it. It is the same as music or art. You have to have the fundamental flare.

Q: You mean this is almost hereditary?

T: It is not hereditary. I would not say it is hereditary. But it is an inborn characteristic. It is like writing poetry. There are sports of nature; they occur, we do not know why they occur, we do not know why they occur in a particular family or at a particular time. But unless the man has the innate ability and liking to do mathematics — and, mind you, it is a self-absorbing and completely absorbing study — he can work his head off and he can make a progress up to a point, but he will never be outstanding.

Figure 1. The platform party before the closure of the ceremonies of the 1965 Convocation, May 28, 1965. Earle Wilcox Crampton and Herbert Tate in the foreground. Source: *The McGill News*/McGill University Archives, PR012236.

9. Closing Comments

Q: Professor Tate, I'm just gonna get personal for a moment. I know most faculty people have sort of given their lives over to teaching, and teaching has been their main hobby, their main profession. So I wondered whether you have had any other interests over the years.

T: Well, my main interests have always been teaching and my students. I was a cricketer at one time. I was a footballer at one time. But that's a long, long time ago. That's before I came to McGill.

Q: Now what are your plans immediately? You are still teaching, you have been on post-retirement teaching, and I wondered what you saw in the coming years.

T: At the present moment, I have no definite plans. I vaguely hope to do a lot of things, which I meant to do over the years and did not have time to do.

Q: Thank you very much, Professor Tate.

Acknowledgments

Thanks are due to my friend David Bellhouse and my wife Johanna Nešlehová for encouraging me to complete this project. The technical support of McGill University archivist Jean-Marc Tremblay and the assistance of Lucie Čermáková in transcribing the interview are also gratefully acknowledged.

References

Claudian, V. & Tate, H. (1937). Problems and Solutions: Elementary Problems: Solutions: E248. *Amer. Math. Monthly*, 44 (6), 390–391.

Gelbart, A. & Tate, H. (1936). Problems and Solutions: Elementary Problems: Solutions: E194. *Amer. Math. Monthly*, 43 (8), 497.

Levenson, M. E., Locke, J. F. & Tate, H. (1938). Problems and Solutions: Advanced Problems: Solutions: 3766. *Amer. Math. Monthly*, 45 (1), 56–58.

Musselman, J. R., Tate, H. & Sparrow, C. M. (1938). Problems and Solutions: Advanced Problems: Solutions: 3807. *Amer. Math. Monthly*, 45 (10), 700–702.

Tate, H. (1929). *Interest, Annuities and Bonds*. Sir Isaac Pitman & Sons, Toronto, Canada.

Tate, H. (1937). *Elements of the Mathematical Theory of Interest*. Sir Isaac Pitman & Sons, Toronto, Canada.

Tate, H. (1946). *Elementary Mathematical Analysis*. Sir Isaac Pitman & Sons, Toronto, Canada.

Tate, H. (1947). *Mathematical Theory of Interest*. Sir Isaac Pitman & Sons, Toronto, Canada.

Underwood, R. S. & Tate, H. (1938). Problems and Solutions: Advanced Problems: Solutions: 3794. *Amer. Math. Monthly*, 45 (6), 393–394.

Mathematical Creativity: The Unexpected Links

Amine El-Sahili[1], Nour Al-Sharif[2], Sahar Khanafer[3]

Lebanese University, Beirut, Lebanon

Abstract

Creativity in mathematics is identified in many forms or we can say is made up of many components. One of these components is *The Unexpected Links* where one tries to solve a mathematical problem in a nontraditional manner that requires the formation of hidden bridges between distinct mathematical domains or even between seemingly far ideas within the same domain. In this article, we design problems that express unexpected links in mathematics and suit students of intermediate and secondary levels. We prove their feasibility through teachers' testimonies and through introducing them in classrooms and collecting students' attitudes with respect to understanding and interest. Results confirm that students can sense such component and that designed problems had caught teachers' and students' interest.

Keywords: *Creativity, Mathematical creativity, Unexpected Links, Classroom problems.*

1 Introduction

In 2004, Sriraman conducted a qualitative study in which he interviewed five creative mathematicians to get an insight on some characteristics of mathematical creativity. He ended his study with an inspiring conjecture that captured our interest:

---

[1] *Email:* *sahili@ul.edu.lb*

[2] *Email:* *nour.alsharif.87@gmail.com*

[3] *Email:* *saharkhanafer90@gmail.com*

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 417 – 464

"It is my conjecture that in order for mathematical creativity to manifest itself in the classroom, students should be given the opportunity to tackle non-routine problems with complexity and structure - problems which require not only motivation and persistence but also considerable reflection." (Sriraman, 2004)

We believe that problems demonstrating unexpected links between two or more domains in mathematics represent one type of the non-routine problems addressed by Sriraman in his conjecture. Through well designed problems of this kind, students could be trained to find unusual connections between seemingly far domains in mathematics in order to reach a solution.

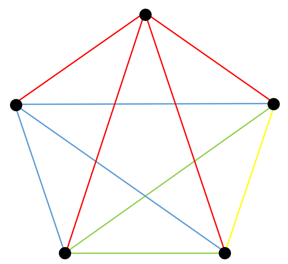
This approach in solving a mathematical problem is viewed in the work of many creative mathematicians. In 1896, Hadamard (Hadamard, 1893; Hadamard, 1896) and De la Vallée Poussin (De la Vallée Poussin, 1896) established and proved (independently) the famous *Prime Number Theorem* using complex analysis. This theorem describes the general distribution of prime numbers among positive integers. It states: *If* $\pi(x)$ *is the number of primes less than or equal to* x, *then* $\lim_{x\to\infty}\frac{\pi(x)\ln(x)}{x}=1$*; that is,* $\pi(x)$ *is asymptotically equal to* $\frac{x}{\ln(x)}$ *as* $x\to\infty$. Although there is no clear connection between complex analysis and the distribution of prime numbers, the proof depends greatly on Riemann's zeta function from complex analysis.

Another example is related to the infinity of primes. Fürstenberg defined a topology on the set of integers $\mathbb{Z}$ and linked it with elementary properties on numbers to prove that the set of prime numbers is infinite (p:5) (Aigner & Ziegler, 2010).

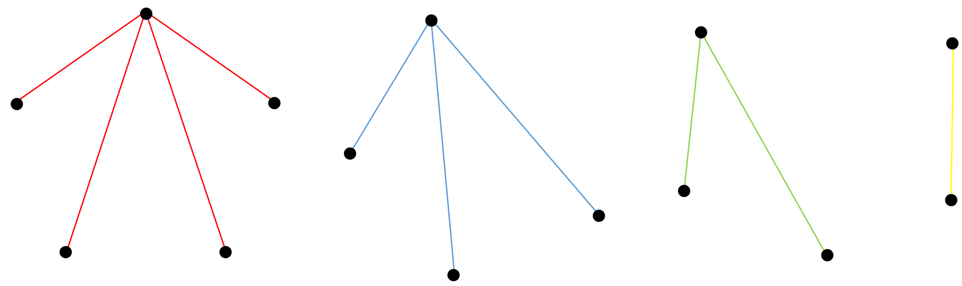
Moreover, in graph theory, Erdös and Rényi introduced for the first time the probabilistic methods to prove the existence of some graphs that are usually difficult to find (Aigner & Ziegler, 2010). The usage of this method represents the unexpected link between graph theory and probability.

In addition, the friendship theorem is a real situation problem that was translated into a graph theoretical problem and then solved using both graph theory and linear algebra techniques. It states: *"Suppose in a group of people we* *have the situation that any pair of persons has precisely one common friend. Then there is always a person who is everybody's friend."*

Also, in graph theory, Tverberg (1982) established a proof of a theorem about the decomposition of a complete graph into complete bipartite graphs. This proof makes use of a system of linear equations to show that the minimum number of stars necessary to cover a complete graph is $n-1$. The following example illustrates this theorem for the complete graph K_5:



The decomposition of K_5 into the four stars $K_{1,4}, K_{1,3}, K_{1,2}$ and $K_{1,1}$ respectively:



After reviewing the above examples, one might think that such approach to solve mathematical problems can only be used by professional mathematicians or postgraduate math students since they possess deep knowledge in mathematics, yet we claim that such approach can be inserted in the educational curriculum of intermediate and secondary students. In this way students will be able to experience such type of creativity in solving mathematical problems by constructing links among distinct domains. This is clearly shown in the content of this paper where examples are given and tested to prove the credibility of this claim.

2 Literature Review

The examples mentioned above, and others, provided us with the very first flame that ignited the idea that there is a correspondence between solving mathematical problems using unexpected links and mathematical creativity.

2.1 Mathematical Creativity

For a long period of time, the dominant view was that creativity in mathematics is limited to “genius” individuals (p.148) (Weisberg, 1988). This view has shifted and, today, mathematical creativity is seen as a skill that can either be fostered and encouraged or suppressed and deprived (Silver, 1997; Leikin, 2009). Due to this new view, many contemporary research in mathematics education aim to define mathematical creativity, identify its characteristics, search for tools that assess it, or even tools that implement it in the general school population. A common point in all of these researches is that there is no universally accepted definition of mathematical creativity or even creativity in general (Sriraman, Havold, and Lee, 2013; Treffinger et al., 2002; Mann, 2005; Haylock, 1997).

In this paper, we adopt the definition that was suggested by Sriraman during a conversation with Liljedahl (Liljedahl & Sriraman, 2006). He differentiated between mathematical creativity at two levels:

- At the professional level, mathematical creativity can be defined as the:
    a) Ability to produce original work that significantly extends the body of knowledge (which could also include significant synthesis and extension of ideas).
    b) Ability to open avenues of new questions for other mathematicians.
- At the classroom level, mathematical creativity is defined as the process that results in:
    a) Novel and/or insightful solutions.
    b) The formulation of new questions and/or possibilities that allow an old problem to be regarded from a new point of view.

2.2 Unexpected Links

Throughout the literature, many are the ideas that indirectly support our claim. Sriraman, in his conversation with Liljedahl (Liljedahl & Sriraman, 2006), gave an example to clarify what he meant by saying: “original work…could also include significant synthesis and extension of ideas” while defining mathematical creativity at the professional level. The example was Witt's proof (1931) of Wedderburn's theorem; *a finite division ring is a field*, which uses algebra, complex analysis and number theory.

It is well known that a commutative division ring is a field, so it is brilliant how the finite condition on the division ring forces the multiplication in this ring to become commutative and thus transforms it into a field. In this subject, Herstein writes *“It is so unexpectedly interrelating two seemingly unrelated things, the number of elements in a certain algebraic system and the multiplication of that system”* (as cited in Aigner & Ziegler (2010), p.31). Also Aigner and Ziegler, in their book *Proofs from the Book* (p.31), comment that this proof *“combines two elementary ideas towards a glorious finish”*. Thus the combination of these different domains in mathematics to formulate a proof of the theorem is considered to be mathematically creative.

Booden states that “sometimes creativity is the *combination of familiar ideas in unfamiliar ways*” (p: xi) (Booden, 2004). Chamberlin and Moon note that mathematical creativity is realized when one *creates a non-standard solution of a problem* that can be solved using standard methods (Chamberlin & Moon, 2005). Nadjafikhah et al. conclude their study with three points that could consist a creative act in mathematics: creating a new fruitful mathematical concept, *discovering an unknown relation*, and recognizing the structure of a mathematical theory (Nadjafikhaha, Yaftianb, & Bakhshalizadehc, 2012).

Moreover, in almost all contemporary research about fostering or assessing mathematical creativity, experiments were evaluated depending on the three components of creativity that were originally defined by Torrance (1966, 1974) and then redefined to suite the assessment of creativity in mathematics. One of them is flexibility: “flexibility refers to apparent *shifts in approaches* taken when generating responses to a prompt” (Silver, 1997). This is consistent with the notion of *unexpected links* where both tend to solve a problem in an unusual approach that relates several distinct domains in mathematics.

In an article about aesthetics and creativity, Brinkmann and Sriraman sent a questionnaire to some mathematicians (Brinkmann & Sriraman, 2009). Some of the answers assert that mathematicians sense elegance and beauty in proofs that demonstrate unexpected links between two or more domains in mathematics:

- "A particularly intensive appeal comes from the suddenly (and sometimes unexpected) pure discovery, the clear understanding of a mathematical phenomenon, often from a completely new perspective, a *new harmonic interplay of different fields, first appearing not to be related to each other*."
- "Beauty within mathematics manifests itself on the one hand by typical mathematical-logical arguments, especially if these arguments show *unexpected and important connections, in an (at first) surprising manner and then mostly also in a surprising simple manner*."

They conclude that aesthetic appeal plays a crucial role in the creative work of contemporary mathematicians. And since there is a governing call to view school students as budding mathematicians, *it is ironic that aesthetics has not received much attention by the community of mathematics educators* (Brinkmann and Sriraman, 2009); Especially that in order to motivate students towards getting engaged in creative mathematical thinking or even mathematical thinking, they must sense the beauty of mathematics (Sinclair, 2009). For as Hardy (1940) said: "*There is no permanent place in the world for ugly mathematics*" (Hardy, 1940). Thus unexpected links is one method that can motivate students toward beautiful creative mathematical thinking.

As reinforcement to our claim, we testify by Poincaré's statement:
"Elegance may result from the feeling of surprise caused by the unlooked-for occurrence of objects not habitually associated. In this, again, it is fruitful, since it discloses thus relations that were until then unrecognized. ***Mathematics is the art of giving the same names to different things****"* (Verhulst, 2012).

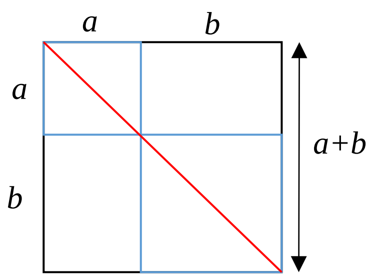
2.3 Calls to Foster Mathematical Creativity

While reviewing the general literature about mathematical creativity, we stood upon many recommendations to foster it in classrooms. Sriraman emphasized that: “it's in the best interest of the field of mathematics education that we identify and nurture creative talent in the mathematics classroom” (Sriraman, 2004). In fact, limiting the use of mathematical creativity in classrooms transforms mathematics into a set of skills to master and rules to memorize (Mann, 2005), whereas “the wellsprings of mathematics are not utility and relevance, but creativity, imagination and appreciation of the beauty of the subject” (Whitcombe, 1988).

Furthermore, researchers recommend mathematics educators to identify and develop mathematical creativity (Nadjafikhaha, Yaftianb, & Bakhshalizadehc, 2012). “All students, especially those with potential talent in mathematics, need academic rigor and challenge as well as creative opportunities to explore the nature of mathematics and to employ the skills they have developed” (Mann, 2005).

3 Unexpected Links: Among and Within Branches of Mathematics

First algebraists established some algebraic formulas through geometric demonstrations. Euclid, in his book *Elements*, uses geometric notions to express what we today consider as algebraic formulas. For example, he states and illustrates the following proposition: *“If a straight-line is cut at random, then the square on the whole (straight-line), is equal to the (sum of the) squares on the pieces (of the straight line), and twice the rectangle contained by the pieces”* (p.52-53) (Fitzpatrick, 2007).



The algebraic translation of this proposition presents the identity: $(a+b)^2 = a^2 + 2ab + b^2$ which is gulped into students' memory at the intermediate levels with just a simple proof by expansion. Whereas if demonstrated as it was originally developed by Euclid, it will certainly capture their attention.

Moreover, Al-Khawarizmi, in a printed translated copy of his book (Rosen, 1831), provided the geometric illustration of the equation $x^2 + 10x = 39$ through which he found one of its solutions, $x = 3$. His method can be generalized to find one solution, $x = \frac{-p + 4\sqrt{4q + p^2}}{2}$, of any quadratic equation of the form $x^2 + px + q = 0$.

In addition, a simple demonstration of unexpected links, in algebra, grows upon understanding the nature of "equal". At early stages, students might enclose on the idea that equality is a trivial notion, for example: $2 = 2$ or $1 \neq 2$. Whereas, while ascending to higher levels of education, they encounter a much more sophisticated understanding of equality. For example, after being introduced to fractions, students will view the equality of $\frac{8}{4} = \frac{6}{3} = 2$ or $\frac{513}{456} = \frac{9}{8}$. In intermediate levels, they get familiarized with more complex equalities such as $\sqrt{2} + \sqrt{5} = \sqrt{7 + 2\sqrt{10}}$ which is an expression of equality among unequal objects at the first glance. At higher levels, students will counter an advanced expression of equality and inequality as $\sqrt{2} \neq \frac{p}{q}$ with $p, q \in \mathrm{N}$. Another example is that concerning the representation of a real number as a combination of imaginary numbers, such as $3 = \left(1 + \sqrt{-5}\right) + \left(2 - \sqrt{-5}\right)$ or $5 = \left(2 + \sqrt{-1}\right)\left(2 - \sqrt{-1}\right)$.

Moving to the domain of geometry, a notable observation in Euclid's Elements is the interrelated study of two different geometric structures, circles and triangles. For example, Euclid made use of circles "to construct an equilateral triangle on a given finite straight-line" (Euclid's st1proposition, book I), (Fitzpatrick, 2007) (p.8).

Furthermore, constructing precise structures (shapes) out of random ones is one of the beams of hidden links within plane geometry. One of these shapes is Euler line that joins the orthocenter, centroid, and circumcenter of any triangle. Another shape is the *Nine-Point Circle;* the feet of the three altitudes of any triangle, the midpoints of the three sides, and the midpoints of the segments from the three vertices to the orthocenter, all lie on this circle. Surprisingly, the center of this circle lies on Euler line.

On the other hand, it is remarkable and interesting to shed the light on a theorem that links algebra and projective geometry. It states: *in a finite projective plane, Desargues' theorem implies Pappus' theorem follows from the algebraic result that a finite division ring is a field*. *"The idea to pursue here is the assignment of number-like objects to the elements (points, lines) of various geometries"* (p.254) (Kleiner, 2012). An algebraic is system is constructed and studying the geometric properties is replaced by studying the associated algebraic system.

Moreover, Kleiner emphasizes the importance of building bridges between distinct domains of mathematics and how such bridges that were built between algebra and geometry resulted in the creation of a new field which is analytic geometry. *"Building bridges between different, seemingly unrelated, areas of mathematics is an important and powerful idea, for it brings to bear the tools of one field in the service of the other"* (p.12) (Kleiner, 2012).

Also, Pythagoras theorem represents the bedrock that resulted in linking arithmetic and geometry and thus in the creation of a new mathematical branch which is analytic geometry. *"The Pythagorean theorem was the first hint of a hidden, deeper relationship between arithmetic and geometry, and it has continued to hold a key position between these two realms throughout the history of mathematics"* (p.2) (Stillwell, 2010).

Another good example shows the link between analysis and graph theory, where Cauchy-Schwarz inequality: $\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$ *for real numbers*

$a_i, b_i,\ 1 \le i \le n$, is applied to get, surprisingly, the following result: *a graph on* n *vertices and without triangles, has at most* $n^2/4$ *edges*.

As a conclusion, we choose to end this part with the following quotation:

"Where things get really interesting is when unexpected bridges emerge between parts of the mathematical world that were previously believed to be very far remote from each other in the natural mental picture that a generation had elaborated. At that point one gets the feeling that a sudden wind has blown out the fog that was hiding parts of a beautiful landscape" (p.3) (Connes).

4 Embedding Unexpected Links through Classroom Examples

In a regular mathematics school curriculum, students are deprived the opportunity of realizing hidden links among different mathematical domains and thus they grow up to believe that such links do not exist.

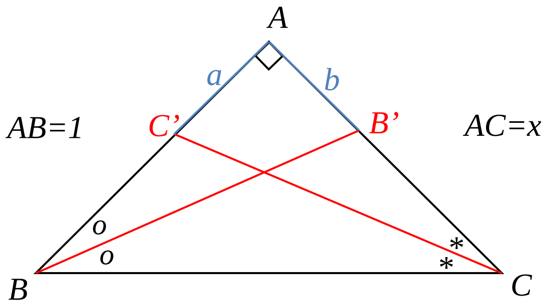
In this section, we propose examples that open the eyes of school students towards the existence of links between mathematical domains that seem very far apart. Note that as we descend from higher levels to lower ones, creating and demonstrating examples of this kind becomes more and more difficult. This is due to the accompanied limitation in students' acquired mathematical knowledge.

4.1 Real Functions and Geometry

This is an example for a third-secondary class in which a geometric problem is solved by the aid of an exponential function:

Problem: Let ABC be a right triangle at A, with (BB') and (CC') the bisectors of $A\hat{B}C$ and $A\hat{C}B$ respectively, such that $BB' = CC'$. Show that if the function $f(t) = \dfrac{e^t - e^{-3t}}{2}$ is strictly increasing, then ABC is isosceles.

A. Suppose $AB = 1$ and $AC = x > 0$. Let $AC' = a$ and let $AB' = b$.



1. Find a and b in terms of x. (Hint: apply the *Angle bisector Theorem*: consider a triangle ABC, if AE is the angle bisector of $B\hat{A}C$ then $\frac{EB}{EC} = \frac{AB}{AC}$).

2. Using the previous part, show that:

$$x + \frac{x}{\left(x+\sqrt{1+x^2}\right)^2} = \frac{1}{x} + \frac{\frac{1}{x}}{\left(\frac{1}{x}+\sqrt{1+\frac{1}{x^2}}\right)^2}$$

(Hint: Show that $a^2 + x^2 = b^2 + 1$)

3. Find a function g such that $g(x) = g\left(\frac{1}{x}\right)$.

B. Consider the hyperbolic sine function $\sinh(t) = \frac{e^t - e^{-t}}{2}$ and the hyperbolic cosine function $\cosh(t) = \frac{e^t + e^{-t}}{2}$.

1. Define the domain of definition of $\sinh$ and $\cosh$.
2. Show that $(\sinh(t))' = \cosh(t)$ and $(\cosh(t))' = \sinh(t)$.
3. Study the variations of $\sinh$ and $\cosh$.
4. Deduce that for $x > 0$ there exists $t > 0$ such that $x = \sinh(t)$.
5. Show that $\cosh^2(t) - \sinh^2(t) = 1$.

C. Let $g(x) = x + \frac{x}{\left(x+\sqrt{1+x^2}\right)^2}$ and $x = \sinh(t)$ with $t > 0$.

1. Find the exponential function $f(t)$ such that $f(t) = g(x)$.

2. Show that f is strictly increasing.

3. Deduce that g is strictly increasing.

D. Using the previous parts, deduce that ABC is isosceles. (In other words, deduce that $x = 1$)

As much as this problem can catch the attention of the students, it will before, catch the attention of their teachers. Even at the level of math teachers, which most of them are expected to be holders of at least a bachelor degree in pure mathematics, the main statement of the problem will be surprising and finding the link will not be an easy task. They will be in need of some steps similar to the ones mentioned above to formulate the big picture of the solution.

4.2 Inequalities and Functions

Here is an algebraic problem that can be introduced to first and second secondary students. It reflects a link between inequalities and variation of functions.

Problem. Compare $a = \sqrt[3]{5 - \sqrt[3]{2}}$ and $b = \sqrt[3]{2 + \sqrt[3]{4}}$.

1. Compare $\sqrt[3]{2}$ and $\dfrac{-1+\sqrt{13}}{2}$.

2. Let $f(x) = x^2 + x - 3$. Show that $f(x) < 0$.

3. Compute $b^3 - a^3$.

4. Deduce that $a > b$.

4.3 1 in Disguise

The following problem demonstrates what we have pointed out previously that unexpected links can appear within the understanding of simple mathematical objects, such as "=". The problem will show that "1" can take a complex surprising disguise: $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}$. The following problem is suitable for second secondary students:

Problem: Show that $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}$ is rational.

1. Show that $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}$ is a root of x^3+3x-4.
2. Prove that $x^3+3x-4=(x-1)(x^2+x+4)$.
3. Show that x^3+3x-4 has one real root and two imaginary roots.
4. Deduce that $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}=1$; that is, $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}$ is rational.

4.4 Triangles and Prime Numbers

For students of grade 10, we suggest this example that links triangles and prime numbers:

Problem: Let ABC be a right triangle at A. Let $A\hat{B}C=x$ and $A\hat{C}B=y$ such that $x>5$ and $x,y\in \mathbb{N}$ (measure in degrees). Show that if x is prime, then y is prime, provided that y is not divisible by 7.

1. Suppose that y is not prime number. Show that y has a divisor less than 10.
2. Deduce that y has a prime divisor p less than 10.
3. What is the set of possibilities of p?
4. Show that if $p\in\{2,3,5\}$, then x is no more prime.
5. Deduce that, according to the given, y is prime.

This solution provides the student with a new way of thinking and solving where s/he doesn't just view x and y as angles that sum up to 90. S/he manipulates these integers in a simple logical manner to reach the result.

4.5 Linked properties in a triangle: Heights, Perpendicular Bisectors and Bisectors

In the 8th grade, students learn that in any triangle, the three heights, the three bisectors, as well as the three perpendicular bisectors, meet in one point. These three concepts are apparently unrelated due to the different nature underlying their definitions. Surprisingly, we can find out how tightly they are linked to each other in the following manner:

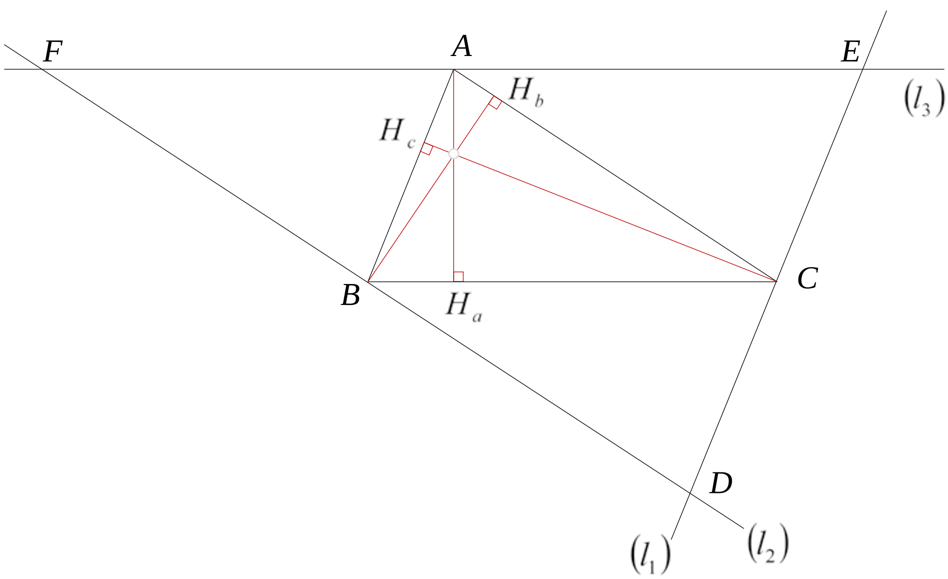
4.5.1 Perpendicular Bisectors and Heights

In the first part, we link a property of the perpendicular bisectors in a triangle to a property related to its heights:

Knowing that the perpendicular bisectors of any triangle meet in one point prove that the three heights also meet in one point?

Consider any triangle ABC. Let AH_a, BH_b and CH_c be the heights relative to the bases BC, AC and AB respectively. Since we don't know about the nature of their intersection, if it is one point or it forms a triangle, we will hide this intersection.

1. Let (l_1) be the line passing in C and parallel to (AB), (l_2) be the line passing in B and parallel to (AC), and (l_3) be the line passing in A and parallel to (BC). Let D, E and F be the points of intersection of (l_1) and (l_2), (l_1) and (l_3), (l_2) and (l_3) respectively.



2. Prove that $ABCE$, $ACBF$ and $ACDB$ are parallelograms.

3. Prove that A, B and C are the midpoints of $[EF]$, $[FD]$ and $[DE]$ respectively.

4. Deduce that (AH_a), (BH_b) and (CH_c) are the perpendicular bisectors of $[EF]$, $[FD]$ and $[DE]$ respectively.

5. Deduce that the 3 heights of triangle ABC meet in a common point.

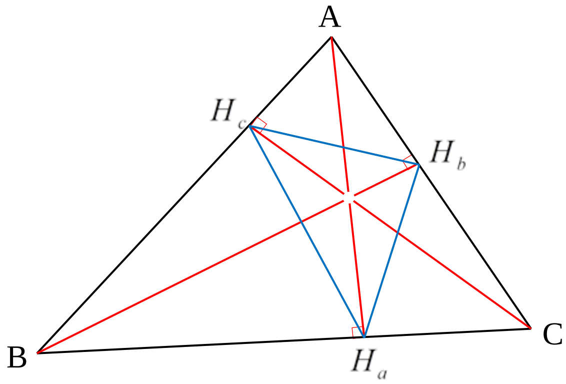
4.5.2 Angle Bisectors and Heights

In this section, another interesting hidden link between bisectors and heights is illustrated.

Knowing that the angle bisectors of any triangle meet in one point prove that the three heights also meet in one point?

Let ABC be any triangle. Let H_a, H_b and H_c be the feet of the heights drawn from A, B and C, respectively, to the relative bases.

.



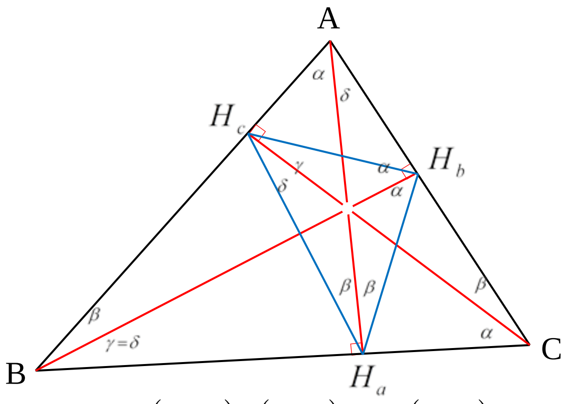
1. Prove that BCH_bH_c is an inscribed/cyclic quadrilateral.

2. Deduce that $H_c\hat{H}_bB = H_c\hat{C}B = \alpha$, $H_b\hat{H}_cC = H_b\hat{B}C = \beta$ and $H_b\hat{C}H_c = H_b\hat{B}H_c = \gamma$.

3. Show that ABH_aH_b is an inscribed quadrilateral.

4. Deduce that $B\hat{A}H_a = B\hat{H}_bH_a = \gamma$, $A\hat{H}_aH_b = \delta$ and $H_a\hat{A}H_b = \beta$.

5. Prove that ACH_aH_c is an inscribed quadrilateral.
6. Deduce that $C\hat{H}_cH_a = \beta$, $H_c\hat{H}_aA = \delta$ and $\alpha = \gamma$.



7. Deduce that the heights (AH_a), (BH_b) and (CH_c) meet in one point.

The teacher can end up this problem by posing some *open questions*:

- How can we show that if the three heights of any triangle meet in one point, then:
    1. The three angle bisectors also meet in one point?
    2. The three perpendicular bisectors also meet in one point?
- Is it possible to find equivalence between both properties?

4.6 Attractive Induced Formula

Now we move into a lower level. We propose the following example for grade 7 (second intermediate). It is based on a well-known proposition in these grades:

$$\text{If } \frac{a}{b} = \frac{c}{d} \text{ then } \frac{a}{b} = \frac{a+c}{b+d} \ .$$

We can represent this formula in a more beautiful and attractive way:

$$\frac{12}{35} = \frac{1212}{3535} = \frac{121212}{353535} = \cdots$$

$$\text{Or:} \qquad \frac{431}{205} = \frac{431431}{205205} = \frac{431431431}{205205205} = \cdots$$

Introducing the above property with the symphonic numerical equality will not reflect any link at the first glance. However, certain demonstrations will clarify this hidden link and can allow students in lower intermediate levels to understand and be interested in such attractive formula.

1. Using your calculator find the ratios $\frac{13}{75}$, $\frac{1313}{7575}$, $\frac{131313}{757575}$.
2. What can you conclude? $\left(\frac{13}{75} = \frac{13131313}{75757575} = \cdots\right)$
3. Can you give a 3-digit example with the same previous feature?
4. Show that: "If $\frac{a}{b} = \frac{c}{d}$ then $\frac{a}{b} = \frac{a+c}{b+d}$ " .
5. Use this property to give the reason behind the above equalities?
6. Can we deduce that $\frac{25}{471} = \frac{2525}{471471}$? Why? (try to use the formula)
7. Apply the strategy on the 3-digit-number example.

We claim that this example will lighten an aesthetic numerical form of the main property. It will attract students, stay in their minds and develop their way of thinking to discover similar properties.

5 Research Methodology

Our notation: *The Unexpected links*, between distinct mathematical domains or even between seemingly far ideas within the same domain, is a component of mathematical creativity that is sensed by students of pure mathematics in postgraduate levels. However, students of lower levels, especially school students, are deprived from the opportunity to experience such beautiful and creative links in mathematics.

In addressing this problem, we hypothesize that *problems that hold unexpected links can be designed to suit intermediate and secondary scholastic levels, where they can be understood by students and can catch their interest*.

With this hypothesis the following questions rise up:

1. Is *The Unexpected Links* a form of mathematical creativity?
2. Are problems that demonstrate the unexpected links applicable in regular mathematics classrooms?
3. Can students understand such problems?

4. Are these problems interesting for students?

The first question was answered through the illustrations we provided from the literature review and history of mathematics. The second question is answered through teachers' interviews and questionnaires. Whereas for the last two questions, their answers are provided through the observations noted while demonstrating the problems in classrooms and through the data collected from post questionnaires distributed to the students.

5.1 Design

Our experiment is divided into two parts, a teachers' experiment and a student’s one. For the teachers' experiment, we were obliged to use two procedures according to the availability of teachers and their time. Sometimes we had the opportunity to group a number of teachers together and introduce the problems to them; while other times, when teachers had no common free time, we had to introduce the problems to each teacher separately.

For the same reasons, we were able to conduct semi-structured interviews on a group of teachers while on others we had to pass a questionnaire. Note that both interviews and questionnaires held the same contents but interviews gave us a better detailed insight on teachers' opinions. For the students' experiment, we introduced each single problem into a classroom. All classrooms were subjected to post questionnaires.

5.2 Participants

We applied our experiment on teachers from two schools and on students from one school. We had the chance to conduct the experiment on 10 mathematics teachers: 4 teachers of only secondary levels, 2 teachers of only intermediate levels, and 4 teachers of both levels. All have been mathematics teachers for more than 7 years except for one intermediate level math teacher whom has been teaching mathematics for only 3 years. Thus all are well experienced in teaching mathematics. As for students, we introduced our problems in 7 classrooms: 4 secondary levels and 3 intermediate. Each classroom contained at least 16 students.

5.3 Data Analysis

We will present our data in 3 sections:

- Teachers' interview: for each problem, we provide teachers' opinions and then we draw a conclusion.
- Students' experiment: for each problem, we provide the procedure, step by step, by which it was introduced and we provide students' interaction in each step, when remarked. Also we provide some observations noted throughout the whole demonstration.
- Questionnaires' analysis: a post questionnaire was distributed to each student. This questionnaire investigates students' impressions towards the presented problem.

5.3.1 Teachers' Interviews

We denoted the questions by Q1, Q2, Q3…etc. and we named the teachers by T1, T2, T3…etc. The statements provided in square brackets are the researchers' intervention in order to rectify or explain some points to the interviewed teacher. If a teachers' opinion is not provided, then this means that the answer is missing. The questions of the teachers' interview are provided in appendix B.

1) Real Functions and Geometry:

This is a problem we proposed for 3^{rd} secondary students. 7 teachers' opinions are provided here.

Q1 :

T1 You can explain all concepts and make students understand them, but with respect to them this is a very high level problem. They won't understand anything; maybe 1 or 2 students will be able to proceed with you. Now, you can tell them the rules of $\sinh$ and $\sinh$ if they know exponential functions; you can make them understand all concepts as you like. But I think students can't relate all parts together and notice how the question is proceeding… Even if 1 or 2 students proceed with you, they will get lost and will need to review the problem at home.

T2 No, they know all.

T3 No.

T4 No.

T5 No.

T6 Yes, rules, calculation, and geometric properties that are rarely used.

T7 No.

Q2 :

T1 No, a part alone no. But you will need to work with clever students that can solve. Sometimes, even those things that we find easy (cross multiplication ...etc.) make many students lost. For example, if you ask life science section students to find a in terms of x, only 2 or 3 students can solve it. Now for general science sections, yes maybe more than half of the class can solve such things. If you are addressing general sciences students, yes every part alone is good.

T2 If for general sciences students, then of course all parts are not hard. For life science students, some parts are hard.

T3 No.

T4 No.

T5 No.

T6 Yes, the linkage of geometric exercises to tough irrational functions and composite functions.

T7 No.

Q3 :

T1 Yes, it is hard. If the problem is for general sciences students and the teacher explains every part, then yes students can understand. [This is what I meant, if the teacher explains every part, can they understand it?] Yes they can understand but I am telling you that not all students of general sciences can proceed with you, the clever ones only. [That is the weak students cannot understand?] No, the weak ones cannot understand.

T3 No.

T4 No.

T5 No.

T6 Yes, the final linkage and the use of bijective functions are hard for students to understand. The meaning of $g(x) = g\left(\frac{1}{x}\right)$.

T7 No.

Q4 All teachers responded with *yes*. T1 added *not only students also teachers. I am sure of that, I saw such kinds of examples but this is a very nice one.* Also, T2 added *students will enjoy*.

Q5 All teachers responded with *yes*.

Q6 :

T1 One remark only, that this is a hard question for secondary levels. But as an idea it is very nice and it is very beautiful to link a function in this manner and show that math is interrelated. It is a very nice idea. Maybe if simplified more, the same ideas but in a low dosage, in a way that if introduced into a general sciences section, the whole class will be enjoying it, in my opinion it would be more beautiful.

T2 Really, it is a nice idea.

T6 You can make the hidden link with what we call "verbal questions"; the questions are exactly those of linkage… I was shocked by this function, $f(t) = \frac{e^t - e^{-3t}}{2}$ when I read it at first in the main statement of the problem. [One of the goals is to surprise you and students by the statement of the problem that holds within the linkage of two seemingly far and distinct concepts, and then we go through the proof step by step]. There are other simpler examples on such links; this is my basic note.

T7 Relating the ideas together is hard.

Most teachers agree that the problem makes no use of mathematical concepts that students are not aware of, except for one geometric property, *The Angle-Bisector Theorem*; it is rarely or never used in schools. For the hardness, teachers agree that each part separately is generally not hard, especially for general sciences students, yet the final linkage of the whole problem is hard. They commented that general sciences students can comprehend the solutions, if explained thoroughly, while life science students will face difficulties. The problem was interesting for teachers, specifically the appearance of an exponential function in a geometric problem.

In conclusion, by considering teachers' opinions, the problem is suitable for general sciences sections, provided that all hints we proposed are introduced. In addition, the problem conveys our notation and has the potentiality to fulfil our aim.

2) Inequalities and Variation of Functions:

We proposed this problem for 2^{nd} secondary students. 8 teachers' opinions are provided here.

Q1 All teachers responded with *no*.

Q2 All teachers responded with *no*, except for T8 who responded with *yes*, *the linkage between the function and comparison in part 3 is hard for students to know and deduce alone*.

Q3 All teachers responded with *no*, except for T2 whose answer is missing. Also, T6 added the following comment *it just needs a rearrangement so that they will understand it better*.

Q4 All teachers responded with *yes*, except for T2 whose answer is missing.

Q5 All teachers responded with *yes*, except for T2 whose answer is missing.

Q6 :

T1 It is very nice.

T6 Rearrange the parts to make the concept closer to the level of the class. In rearranging this problem, ask about the study of the function before the comparison of the numbers in part 1. When I firstly read part 1 after the statement of the main problem, I said “why this comparison is put here? What is its role?” This part transmits a sense of weirdness and foreignness and, due to that, I was no more encouraged to solve it. However, when you start by the function, I will feel that it is true and more beautiful. Also the most beautiful is to ask: Compare $\sqrt[3]{2}$ and

$\frac{-1+\sqrt{13}}{2}$ and then put $\sqrt[3]{2}$ on the table. Where is its position? It is between the roots. [That is right, it becomes easier] Aha, you said the word I am looking for! There is a big difficulty in the way you proposed…This linkage is very helpful when it is hard to treat a problem by our familiar methods; it makes the problem easier and useful. But all my remarks don't affect the objective of this research; the objective of the research is sacred and doesn't hold any problem.

T7 The problem is very nice. It is nice how in the problem we move and relate the table of signs with the comparison.

T8 The linkage between the function and comparison in part 3 is hard for students to know and deduce alone. [If I solved with the students to reach this part, is it hard to see that $b^3 - a^3 = f\left(\sqrt[3]{2}\right)$?] Look, this type of questions is for special persons, not for all students. [No, my aim is that all students understand such questions.] Aha, you mean that such problem is given directly to the student? Or he should have been trained on similar patterns? [No, there is no previous training at this level. What is important to me is that the student knows all the concepts in the problem, as studying a function with its table of variations…, and that he understands all the parts to reach the link. I will explain the link for him; could he then understand it all?] Yes, of course. But if you bring the question with its parts and ask a student to solve it alone, there will be parts he can surely solve, but linking the ideas together with such given, having cubic roots, will be a bit hard for him. I am saying this according to what we teach and the examples of functions that we solve in classrooms; they almost contain a square root in a square root. Usually, solving with cubic roots is less familiar than solving with square roots. For this, seeing a square root in the question and dealing with it is very normal, it is “nothing” with respect to students. However, when a cubic root appears, they take more time thinking of it.

As a problem, it is very beautiful. Some people may not face any problem in solving it and they know all its concepts… Students are familiar with such patterns, while in your problem they may feel awkward with the cubic root. But when you explain it and give them the indications, they can understand, but not all students can solve it, alone, from the first time. With respect to the concepts, nothing is new; it is all clear and understood.

All teachers gave positive attitudes towards this problem with respect to used concepts, hardness, and understanding. Some teachers found that linking the ideas together might be hard for students to reach, if they solved alone. So they recommended a simple rearrangement in the structure of the problem. In all, according to teachers, the problem is suitable to fulfill our aim.

3) 1 in Disguise:

This is a problem we proposed for 2^{nd} secondary students. 8 teachers' opinions are provided here.

Q1 All teachers responded with *no*, except for T2 whose answer is missing. Also T1 added the following comment: *But I repeat that the students must be of good level*.

Q2 All teachers responded with *no*, except for T6 who responded with *yes, calculation in part 1, students hate to write it* and T2 whose answer is missing.

Q3 All teachers responded with *no*, except for T2 whose answer is missing.

Q4 All teachers responded with *yes*, except for T2 whose answer is missing.

Q5 All teachers responded with *yes*, except for T2 whose answer is missing.

Q6 :

T2 This is an ordinary question. Students can solve.

T6 It is true that when one looks at this number with cubic root, he can't believe it is equal to 1!... Nice, nice, nice! It is the example that impressed me the most. Really it is a very excellent question. I answered this questions with total conviction; it is beneficial.

T7 If the calculation of part one is simplified, then the question will be nicer.

T8 Nice question. It is nice and easier than the previous one.

This problem scored total acceptance among teachers. It was very attractive; they only remarked that the calculations of part one might be tiresome for students.

4) Triangles and Prime Numbers:

This is a problem we proposed for 1st secondary students. 3 teachers' opinions are provided here. 1 of the teachers responded to all questions while the others gave a general remark.

Q1 T6 responded with *yes, in every step*.

Q2 T6 responded with *yes, they can't solve it alone*.

Q3 T6 responded with *no, after it is explained, it will be understood*.

Q4 T6 responded with *yes, for sure*.

Q5 T6 responded with *yes*.

Q6 :

T6 All the contents of the problem are not acquired in the Lebanese program. This question is useful for grade 10, but the arithmetic is not given in the curriculum. It is a presumptive imaginary question. This kind of mathematics does not exist in our curriculum. It is all new. Parts 1, 2, 3 are understandable but this concept, "suppose not" to reach a contradiction, does not exist in our curriculum. It can be understood. The level of this problem is much higher than the level of 10th grade and it is even not found in the Lebanese curricula.

T7 It is hard. We do not teach such things in school (arithmetic). Even the contradiction is given at the final levels of secondary stages and not in this strong form.

T8 The idea of supposing the contrary rarely appears in the curriculum… Such examples appear in the 10th grade at the end of the year with space. But we are not explaining them anymore. Maybe it is given once through explaining the lesson, but not through the students' applications. This is logic; that is it needs a lot of application and work

> to be understood by students and to be well-established in their minds. If you explain it, the student will understand, but after a week you will see that he forgot it. It is hard for grade 10; you are overburdening the problems!!

Teachers agree that this is a very hard problem for 10^{th} grade students to solve, particularly alone. Yet they don't deny that students can understand the solution if it is well explained.

5) Perpendicular Bisectors, Heights, and Angle Bisectors:

This title holds within two problems that were both proposed for students of grades 8 and 9. The opinions of two teachers are provided here.

Q1 :

T8 Yes, the concept of “cyclic”. We don't use this term, we ask students to “prove that the four points are on the same circle”. The other concepts are all known.

T9 No.

Q2 Both teachers responded with *no*.

Q3 Both teachers responded with *no*.

Q4 Both teachers responded with *yes*.

Q5 Both teachers responded with *yes*.

Q6 :

T8 The student can prove that four points are on the same circle, but we don't use the terms “cyclic” or “inscribed”. In the 8^{th} grade, they learn that four points are on the same circle if they form two right angles facing same diameter, but we don't say it is cyclic since “cyclic quadrilaterals” are not required from them. Also, 8^{th} grade students don't know that *if we have four points forming two equal angles facing the same chord, then the four points are on the same circle*.

The problem of perpendicular bisector and height is suitable for grade 8 ... It is not hard and the student can solve, deduce and be happy and interested with the question. However, the problem of the bisector and height is not suitable for grade 8 since it holds within properties in the circle and the concept of four points on same circle… the circle is given at the end of the year for 8^{th} grade. Thus, it is better to give this problem to students of grade 9 than those of grade 8.

T9 The problem of the perpendicular bisector and height is suitable for all students of grade 8. But that of the bisector and height just suits the special students in grade 8, whereas it can be given for all students of grade 9.

Teachers agree that some terms should be exchanged with more familiar terms in these grades. Also, the perpendicular bisector and heights problem is suitable for 8^{th} and 9^{th} grade students while the bisectors and heights is only suitable for 9^{th} grade students. With such recommendations and modifications, teachers find the problems not hard and understandable.

6) Attractive Induced Formula:

This exercise was proposed for students of grade 7. The opinions of two teachers are provided here.

Q1 Both teachers answered with *yes, the formula: If* $\frac{a}{b}=\frac{c}{d}$ *,then* $\frac{a}{b}=\frac{a+c}{b+d}$.

Q2 Both teachers responded with *yes, it is hard to prove the formula*.

Q3 :

T9 No, if they take the lesson of proportion and variables.

T10 No, if it is all explained in details.

Q4 T9 answered with a *yes*, and T10's response is missing.

Q5 T9 answered with a *yes*, and T10's response is also missing.

Q6 :

T9 A 7^{th} grade special student can understand the formula. Usually we don't give the proportions in grade 7, which is related to this question.…The idea is beautiful and nice to be given to all students.

[In general, if I give this exercise to a 7^{th} grade class and explained it?] If I want to talk in general, I can say that at the age of grade 7 the student can understand this formula if he took the lesson.

T10 Students can understand this formula if it is explained in details. It needs time to be well understood. [Of course it will take a complete period to be presented and clarified]. That is true; it is not easy. And not all the students are going to understand it. If you designed this question for all students, the intelligent one will catch the idea directly, the student who is fair or good enough may also understand it; however it is impossible for a weak student to understand it. I don't expect he can work on this formula; the weak student can study this formula or apply it directly, but he cannot understand the rule and how we moved from the necessary condition $\frac{a}{b}=\frac{c}{d}$ to the sufficient one $\frac{a}{b}=\frac{a+c}{b+d}$. The link is somehow hard for him. The intelligent and the fair students can understand it because in the seventh class they are exposed to similar ideas or even harder ones. Nothing is hard in this question and they can get it part by part.

Teachers agree that students will not be able to comprehend the proof of the formula and some students might find the structure of the formula hard to digest. Yet they agreed that almost all students can understand the formula and the whole problem if explained well.

5.3.2 Students' Experiment

1) Real Functions and Geometry:

We introduced this problem to a 3^{rd} secondary general sciences classroom consisting of 18 students.

1. We stated and provided the drawing of the main problem prompt. While writing the exponential function on board, one student directly commented with astonishment that what brought this function to this problem!
2. We gave students some time to think it out and try to propose how the ideas are related. After a while, they started to comment that what does the

function have to do with the problem? When nobody had an answer, we wanted to proceed, but a student commanded more time to think. Students started to suggest some scenarios, none were correct. Unfortunately, due to lack of sufficient time, we couldn't provide more time to hear more suggestions.

3. We stated part A. We asked students that, with those notations, what are they required to prove? They directly answered that they should prove $x=1$.
4. We stated part A(1). Since we already knew from teachers that students do not know the Angle-Bisector Theorem, we directly provided and demonstrated it. Students were able to comprehend the theorem and apply it to triangle ABC.
5. We asked students to use this theorem and solve part A(1). One student commented that he didn't understand the whole problem, so we re-illustrated. One student said that he found b, his answer was wrong. Due to lack of time, we had to move on and demonstrate the solution. We provided how to find a, then a student directly provided the answer of b. All students understood the steps.
6. As a reminder to be used in later parts, we wrote a and b in terms of x on one corner of the board.
7. We stated part A(2). We told students that to find the solution, they should notice some given from the figure and combine it with the results of the previous part. A student directly pointed out the given which makes use of Pythagoras theorem. Then we demonstrated the solution with the participation of some students.
8. On the same corner of the board, we added the obtained equation.
9. We stated part A(3). Students acknowledged that the function g has the form of the previous equation and thus the answer was illustrated.
10. We commented that, starting from this point, we are relating the geometric problem to functions. We added to the same corner of the board this new result.
11. We asked students if they were able to find the variation of this function g. They responded that they should find g', so we told them to try. When

nobody commented, we told them that finding the variation of g in this manner is very hard and they should follow another path.

12. We illustrated the hyperbolic functions. Directly students commented that they didn't take them in their mathematics classroom. We emphasized that these functions are in terms of exponential functions that they already know.
13. We hastily went through part B, since we had too little time. We asked for the domain of definition of $\sinh$ and $\cosh$, students directly answered correctly. We asked for their derivatives, students were able to suggest how to derive them and gave the correct answers. Then we asked them to deduce the variation of $\sinh$ only, since the next parts depend just on it.
14. We asked: what do we deduce by having $\sinh$ strictly increasing and continuous? Students directly answered that $\sinh$ becomes bijective. Then we remarked that this leads to the following conclusion, part B(3), that: for $x > 0$ there exists $t > 0$ such that $x = \sinh(t)$. We emphasized on this point. Students acknowledged this deduction.
15. We stated part B(5). Students provided the proof of this part.
16. We commented that due to part B(4), we can now apply a change of variable on g in order to find its variation in an indirect way, thus justifying the given in part C.
17. We stated part C(1). Students directly said that they should substitute $x = \sinh(t)$ in $g(x)$. We gave students some time to solve. Due to lack of time, we had to demonstrate before any of the students was able to give an answer. We started solving and students were helping.
18. We stated part C(2). Students were directly able to provide the solution.
19. We stated part C(3). Then, we demonstrated the solution with the assistance of students. On the same corner of the board, we added the following: g is strictly increasing. We asked students to relate our last findings and reach the final result. Students started to suggest, none were correct, until one student said that this gives us that $x = \frac{1}{x}$. Afterwards, we demonstrated how this result is derived and how it leads to $x = 1$.
20. We demonstrated the steps of the whole problem together.

21. We commented that solving the main geometrical problem in geometric methods is really hard, so we went through another simpler path were we linked 2 domains in mathematics to reach the solution.

Most students were participating and anticipating. Some were curious and others were confused.

2) Inequalities and Variation of Functions:

We introduced this problem to a 2^{nd} secondary sciences classroom consisting of 18 students.

1. We stated the main problem. We provided the expansion: $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$ and we gave students time to solve. None gave an answer.
2. We asked students to find the solution through the calculator. All were able to realize that $a > b$ with a minor difference. We commented that although these two numbers are very close to each other, comparing them together in a traditional manner is hard and thus we need to go through another path.
3. We stated part 1. With our assistance, students were able to provide an illustration of the solution.
4. We stated part 2. Students suggested the solution. It was a trivial part for them. We emphasized on the result found here and we asked students whether they can speculate from this point a scenario of the rest of the solution. Some students suggested, but only one had suggested a relatively close scenario where he said to subtract a and b from each other and apply a table of sign.
5. We stated part 3. Students were able to participate in the computation. We asked students if they can link the parts together and find the final solution. Students tried to relate x of part 2 to the answer, but none gave an explicit and clear answer. We had to intervene and help them out.
6. We asked students to try to suggest how the solution will appear throughout the derived results. A student started to give a correct solution. Then, with our aid, students were able to demonstrate the solution.
7. We commented on the situation, that although those two numbers are relatively close to each other, traditional methods couldn't give a

demonstration of the comparison. So, we had to go through another path that involves the usage of a function and its sign.

Students were participating and group discussions were active. Some students expressed interest and anticipation in viewing results while others were not focused.

3) 1 in disguise:

We introduced this problem to a 2^{nd} secondary sciences classroom consisting of 21 students.

1. First, we cleared out what does a rational number mean. Then we stated the main problem. We gave students time to think. Some students used the calculator and found that this number is equal to 1. None of the students was able to find a demonstration of the solution. So we commented that through traditional methods that involve powers of 3, the demonstration might be impossible. Hence we need another untraditional method.
2. We stated part 1. All students suggested that they should substitute the number in the equation and prove it to be zero. Due to lack of time, we had to demonstrate the solution quickly. Students were participating.
3. We stated part 2. Students directly gave a demonstration.
4. We stated part 3. Students didn't know what imaginary numbers were, so we had to modify the question, we asked them to show that x^3+3x-4 has only one real root. Directly students used the previous part and answered that $x=1$ is the real root.
5. We started to connect the parts together by emphasizing on $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}}$ and 1 being real roots of the equation that has only one real root. Then we asked students what does this situation lead to? All answered that they should be equal.
6. We stated part 4. We re-illustrated the whole steps.
7. We commented that in order to solve this problem we had to go through such untraditional path.

Participation and suggesting was a major observation that went through all the problem. While group discussions took place among some students, confusion was noted among others.

4) Triangles and prime numbers:

We performed this experiment in a first secondary class consisting of 24 students.

1. The problem was put on the board in front of students, pointing out on the conditions. At the first glance, all students seemed to be astonished with nothing to say.
2. We illustrated a numerical example of x and y, some students gave other examples, then we showed how it is possible to set all the possibilities of x and y to prove the result, knowing that $1 \le x, y < 90$ and they are integers.
3. We moved to the algebraic proof and started by the supposition "y not prime". All students were surprised and paying attention to this idea. Though they were not familiar with this method of proof, we clarified it in a simple manner and continued.
4. Stating part 1: Students needed a detailed orientation towards this question; we explained the proof and repeated each step to be clear for all.
5. After stating part 2, we gave a numerical example to make the question more understandable. Then, we recalled that every number has prime factors that are found by making its prime factorization. With more explanation, every point was clear till we reached the required result.
6. We stated part 3. Most of the students answered that they are all primes between 1 and 10, and then a student stated the possibilities $\{2,3,5,7\}$.
7. We stated part 4. This is the heaviest part of the problem since it requires deeper thinking and relating the given, definitely by pursuing the path of proving by contradiction! The solution was demonstrated step by step and students shared by agreeing on each step.
8. The final part was to deduce, according to the given, that y is prime. At this level, we rewrote the previous results together till we reach that the possibilities of $p \in \{2,3,5\}$ are rejected. Here, a student knew that the remaining possibility of p is 7. Then, we explained the reason behind rejecting $p = 7$. Afterwards, a student said: "then it means y should be

prime” and another one added: “you tried all possibilities to reach the required result”.

9. We summarized the whole parts and students understood that the main supposition was wrong and so y is prime.

Students revealed some participation when each part was stated and they tried to give suggestions of the solutions. This didn't hide a sense of distraction by some students, and dullness at certain levels, especially when we started “supposing the contrary”. Sarcastic comments were also heard through the demonstration.

5) Heights and Angle Bisectors:

This problem was demonstrated for 25 students of grade 9.

1. We started by recalling the definitions of the height and the angle bisector in a triangle, and asking if there is something common between these definitions. All students said “no”. We asked about a property of the three angle bisectors of any triangle, a student answered that they meet in one point.
2. We showed the proof of this property in details. Students shared in the proof and gave suggestions; many seemed interested, and some were excited to solve by themselves.
3. We stated the main prompt of problem on the board. A student seemed curious to start solving and asked us to draw the figure with heights on the board.
4. Part 1 was introduced. It was clear that students didn't know the meaning of “inscribed”. So, we clarified what it meant and recalled that if two angles facing same chord are equal then the 4 points forming these angles are on the same circle; then, we showed the proof on the figure.
5. We moved to part 2 and stated that, in a circle, two angles facing the same chord are equal. We pointed on the naming and choice of equal angles; that is, to choose equal angles according to their common extremities on the circle. Then we deduced the corresponding equal angles.
6. Students read the next part 3; they were a bit confused. We proceeded by the same previous pattern, providing the students with some orientations to

reach the result. Directly, we moved to finding the inscribed angles required in part 4. Together we shared with the students and proved the equal angles as explained previously; most of them got the method we are working with.

7. At the time we were moving to part 5, a student directly solved it; s/he named the equal right angles facing the same chord, and deduced that the 4 points, A, C, H_a and H_c are on the same circle. Then, students started naming the equal inscribed angles with some help and orientation. Proving $\alpha = \gamma$ made some confusion in determining the names of the corresponding angles; however, it was noted that a student was naming these angles and all was clear.
8. At the end, we stated part 7 which is the deduction of the main statement of the problem. First we emphasized on the given: *knowing that the three bisectors meet in one point*. Following our directions, students noticed that the heights of ABC are the bisectors of $H_aH_bH_c$ and reached the result.

While explaining the problem, a student showed some fear; he was afraid if this was required for tests or their math program. We calmed him and explained our solely goal to solve together a problem, whose concepts are known for them, and see a certain link. Besides, the names of angles, α, β, δ and γ, made some disturbance, as well as the notations of feet of the heights, H_a, H_b and H_c. These notations were uncomfortable for some students and caused confusion.

In summary, students were participating, speculating, suggesting and showing some interest during the whole period. In addition, some signs of distraction, straying and dullness were recorded. Some students were also playing, not concentrating, and not interested at all.

Finally, it is worthy to state the comment of a student who asked whether the problem is applicable when “the heights are outside the triangle”. It was a remarkable question.

6) Heights and Perpendicular Bisectors:

This problem was explained to 25 students of grade 8.

1. At the start, we recalled the definitions of the height and the perpendicular bisector in a triangle, and then we asked about the common between these two concepts. Many answers were detected: both make $90°$, both are straight lines, and both are parallel (if drawn on the same base).
2. We asked about the property of the three perpendicular bisectors of any triangle; they remembered that they meet in one point. Then, we explained the proof and students shared along with our orientation.
3. The main problem was given. We sketched the figure on the board and started with the steps. ABC is an arbitrary triangle with its three corresponding heights whose intersection is erased. We drew the required constructions in the given.
4. We stated part 1. Many students shared in the answer. They showed that, in each one of these quadrilaterals, the opposite sides are parallel.
5. We proceeded with part 2. With some aid and orientation, a student showed that A is the midpoint of $[EF]$. Then, students recognized the used method and some of them participated in the proof of the other midpoints.
6. We stated part 3. Starting with AH_a, a student knew that it is perpendicular to $[EF]$ by noticing alternating angles between parallel lines. After giving the complete demonstration of AH_a, students continued the proof of the other perpendicular bisectors following our directions.
7. Before reaching the end, we ran out of time and the class was out of control! We were obliged to state the result and explain it without giving them time to think of it. We showed the link by which the heights of triangle ABC became the perpendicular bisectors of triangle DEF, and thus they meet in one point, by using the main given.

There was participation in the class, as well as speculation and suggestions with some group discussions. Some distractions were also detected, and many students were not participating; some were playing and eating as if they had a recess time! Unfortunately, some students were naughty and added sarcastic comments.

7) Attractive Induced Formula:

We proposed this problem to 16 students of grade 7.

1. We started by the first part. Students calculated each fraction and found them equal. One of the students said that "they will be equal since we are repeating the same number up and down". We asked about the reason, then we gave a numerical example that proves to this student that his claim is wrong. A remarkable answer was given by another student: "they are not equal because '5' is one digit, not two". He noticed the remark and almost understood the procedure. We agreed with him/her, and moved to continue the exercise.
2. We stated part 2. After some orientation, a student answered: "every time we repeat a number more times we get the same result". We asked him/her for an example and S/he gave: $\frac{13131313}{75757575}$. Students checked that $\frac{13}{75}=\frac{13131313}{75757575}$.
3. Then, we stated part 3. A student found an example: $\frac{143}{734}=\frac{143143}{734734}$ and found them equal. Another student gave another example: $\frac{125}{137}=\frac{125125}{137137}$.
4. We moved to part 4. Students were confused, they didn't know the meaning of this proposition and weren't that much familiar with variables. There was confusion in making manipulations of variables and changing the signs when moving to different sides of the equation. We demonstrated the parts of the proof and repeated it. But when we kept sensing their confusion, we skipped it and moved to its application.
5. We stated part 5. Due to students confusion, we showed the link directly by putting $a=13$, $b=75$, $c=1300$ and $d=7500$.We applied the formula: $\frac{13}{75}=\frac{1300}{7500}$, then $\frac{13}{75}=\frac{1300+13}{7500+75}=\frac{1313}{7575}$. They participated in reaching it and some said that they "now understand it". We told them that the aim of this property is to apply it on any number we want. Then, we asked them to continue the next step. They did it: $\frac{1313}{7575}=\frac{131300+13}{757500+75}=\frac{131313}{757575}$.

6. Finally, we asked about part 7. We gave an example and showed how to add 3 zeros: $\frac{132}{564}=\frac{132000+132}{564000+564}=\frac{132132}{564564}$. Some students shared and gave correct ideas, and others commented and deduced that “we add zeros according to the number of digits”.

Note that although we didn't have enough time to demonstrate that “the equality of such fractions appears only if the two numbers in the numerator and the denominator have the same number of digits”, students caught this idea from the beginning when we asked them to check it through a numerical example, but we hadn't the opportunity to clarify it by using the formula.

Students of this class recorded a good participation in each step; they were interested, curious, anticipating, speculating, suggesting and showed persistence in finding the solutions. These bright-sided observations were penetrated by some distraction and dullness while explaining the proof of the formula.

5.3.3 Questionnaires’ Analysis

The structure of the post questionnaire is provided in appendix B.

Real Functions and Geometry:

61.1% of the students understood the whole problem and 38.9% understood parts of the problem. 88.9% of the students found the problem interesting.

Inequalities and Functions:

77.8% of the students understood the whole problem and 22.2% understood parts of the problem. 88.9% of the students found the problem interesting.

1 in disguise:

94.7% of the students understood the whole problem and 5.3% understood nothing in the problem (two answers were recorded as missing). 89.5% of the students found the problem interesting (two answers were recorded as missing).

Triangles and Prime Numbers:

45.8% of the students understood the whole problem, 8.3% understood parts of the problem and 45.8% understood nothing in the problem. 62.5% of the students found the problem interesting.

Heights and Angle Bisectors:

43.5% of the students understood the whole problem, 30.4% understood parts of the problem and 26.1% understood nothing in the problem (two answers were recorded as missing). 80% of the students found the problem interesting.

Heights and Perpendicular Bisectors:

56% of the students understood the whole problem, 32% understood parts of the problem and 12% understood nothing in the problem. 80% of the students found the problem interesting.

Attractive Induced Formula:

18.8% of the students understood the whole problem, 62.5% understood parts of the problem and 18.8% understood nothing in the problem. 93.8% of the students found the problem interesting.

6 Conclusions, Limitations and Implications

6.1 Conclusion

According to the analyzed data, "The Unexpected Links" is a component of mathematical creativity that is not patent of professional mathematicians or postgraduate mathematics students. It can be made available to intermediate and secondary school students through well-structured and suitable mathematics problems that meet their acquired mathematical knowledge.

The problems we proposed fulfilled their role, they were understandable for almost all students and they caught students' interest.

According to teachers' testimonies, classrooms' observations and students' interactions, some modifications are required to some problems. We suggest the following:

- A retest of the problem of triangles and prime numbers in higher levels (grades 11 and 12).
- The elimination of the proof of the formula in the attractive induced formula problem and the re-introduction of it in grade 8 where students are more familiar with variables and their manipulations.

- The rearrangement of the parts of the problem of inequalities and functions in accordance with a teacher's recommendation.
- The elimination of the proofs of the intersection of angle bisectors and perpendicular bisectors, since they are distracting the students from the main problem.
- Changing the letters used to denote the angles in the problems concerning heights, angle bisectors and perpendicular bisectors.
- The modification of part 3 of the 1 in disguise problem in a way that doesn't include imaginary numbers.

Appendix A provides the modified structures of the problems.

6.2 Limitations

We can divide the limitations into two kinds: the first in the phase of preparing problems and the second in the experimentation phase. With respect to problem design, as we move to lower level, such kinds of problems are hard to create. This is due to students' limited acquired knowledge. With respect to the experimentation phase:

- We faced some difficulties in finding schools that agree to undergo such experiment.
- We had a difficult time in finding teachers; some barely gave us 5 minute to introduce the problems and interview or apply a questionnaire on them.
- School's administration wouldn't allow us to take more than one period (50-55 minutes) in one classroom.

6.3 Implications

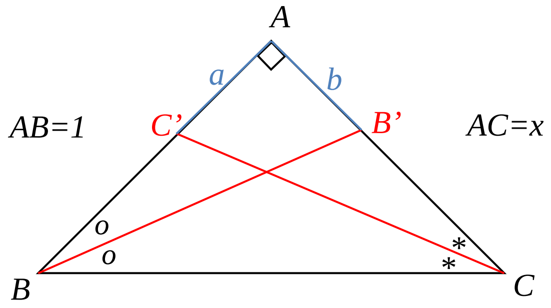
This paper provides evidence that problems demonstrating the unexpected links can be introduced in regular mathematics classrooms; students can comprehend such problems that contain deep links among and within mathematical domains. Now, a single problem has no power to impose a striking impact on students' methods and strategies in solving mathematical problems. Whereas, we believe, even we conjecture that a long term exposition of students to such problems may train them to use this creative method in solving mathematics.

Appendix A

1. Real Functions and Geometry

Let ABC be a right triangle at A, with (BB') and (CC') the bisectors of $A\hat{B}C$ and $A\hat{C}B$ respectively, such that $BB' = CC'$. Show that if the function $f(t) = \frac{e^t - e^{-3t}}{2}$ is strictly increasing, then ABC is isosceles.

A. Suppose $AB = 1$ and $AC = x > 0$. Let $AC' = a$ and let $AB' = b$.



1. Find a and b in terms of x. (Use *the Angle-Bisector Theorem*: consider a triangle ABC, if AE is the angle bisector of $B\hat{A}C$ then $\frac{EB}{EC} = \frac{AB}{AC}$).
2. Using the previous part, show that:

$$x + \frac{x}{\left(x + \sqrt{1+x^2}\right)^2} = \frac{1}{x} + \frac{\frac{1}{x}}{\left(\frac{1}{x} + \sqrt{1 + \frac{1}{x^2}}\right)^2}.$$

(Hint: Show that $a^2 + x^2 = b^2 + 1$)

3. Find a function g such that $g(x) = g\left(\frac{1}{x}\right)$.

B. Consider the **hyperbolic sine function** $\sinh(t) = \frac{e^t - e^{-t}}{2}$ and the **hyperbolic cosine function** $\cosh(t) = \frac{e^t + e^{-t}}{2}$.

1. Define the domain of definition of $\sinh$ and of $\cosh$.
2. Show that $(\sinh(t))' = \cosh(t)$ and $(\cosh(t))' = \sinh(t)$.

3. Study the variations of $\sinh$ and of $\cosh$.
4. Deduce that for $x > 0$ there exists $t > 0$ such that $x = \sinh(t)$.
5. Show that $\cosh^2(t) - \sinh^2(t) = 1$.

C. Let $g(x) = x + \dfrac{x}{\left(x + \sqrt{1 + x^2}\right)^2}$ and $x = \sinh(t)$ with $t > 0$.

1. Find the **exponential function** $f(t)$ such that $f(t) = g(x)$.
2. Show that f is strictly increasing.
3. Deduce that g is strictly increasing.

D. Using the previous parts, deduce that ABC is isosceles. (In other words, deduce that $x = 1$)

2. Inequalities and Functions

Compare $a = \sqrt[3]{5 - \sqrt[3]{2}}$ and $b = \sqrt[3]{2 + \sqrt[3]{4}}$.

1. Let $f(x) = x^2 + x - 3$. For what values of x, $f(x) < 0$?
2. Compare $\sqrt[3]{2}$ and $\dfrac{-1 + \sqrt{13}}{2}$.
3. Compute $b^3 - a^3$.
4. Deduce that $a > b$.

3. 1 in Disguise

Show that $\sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}}$ is a rational.

1. Show that $\sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}}$ is a root of $x^3 + 3x - 4$.
2. Prove that $x^3 + 3x - 4 = (x - 1)(x^2 + x + 4)$.
3. Show that $x^3 + 3x - 4$ has only one real root.
4. Deduce that $\sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}} = 1$; that is, $\sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}}$ is rational.

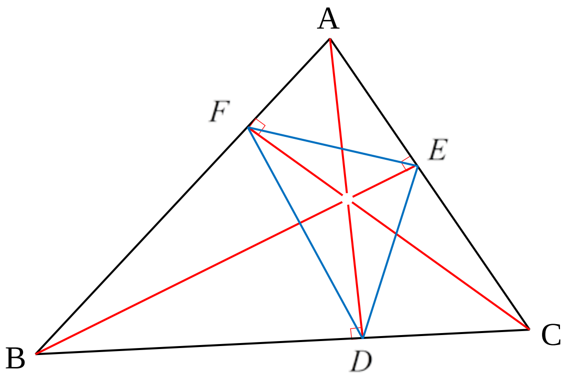
4. Triangles and Prime Numbers

Let ABC be a right triangle at A. Let $A\hat{B}C = x$ and $A\hat{C}B = y$ such that $x > 5$ and $x, y \in \mathrm{N}$ (measure in degrees). Show that if x is prime, then y is prime, provided that y is not divisible by 7.

1. Suppose that y is not prime. Show that y has a divisor less than 10.
2. Deduce that y has a prime divisor p less than 10.
3. What is the set of possibilities of p?
4. Show that if $p \in \{2,3,5\}$, then x is no more prime.
5. Deduce that, according to the given, y is prime.

5. Heights and Angle Bisectors

Let ABC be any triangle. Let D, E and F be the feet of the heights drawn from A, B and C, respectively, to the relative bases. We will hide their intersection. Consider the triangle DEF.



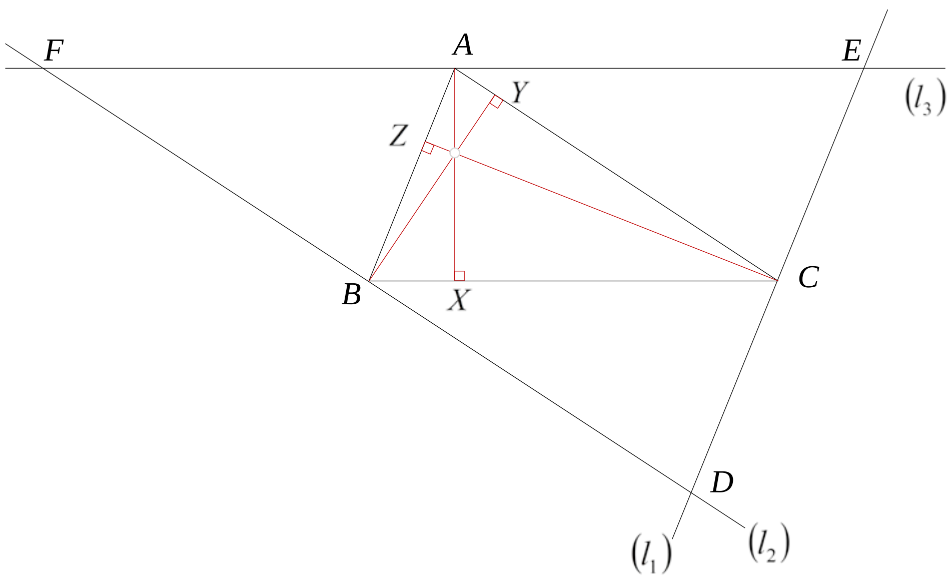
1. Prove that B, C, E and F are on the same circle.
2. Deduce that $F\hat{E}B = F\hat{C}B = x$, $E\hat{F}C = E\hat{B}C = y$ and $E\hat{C}F = E\hat{B}F = z$.
3. Prove that A, B, D and E are on the same circle.
4. Deduce that $B\hat{A}D = B\hat{E}D = t$, $A\hat{D}E = z$ and $D\hat{A}E = y$.
5. Prove that A, C, D and F are on the same circle.
6. Deduce that $C\hat{F}D = y$, $F\hat{D}A = z$ and $x = t$.
7. Deduce that the heights (AD), (BE) and (CF) meet in one point.

6. Heights and Perpendicular bisectors

Consider any triangle ABC. Let AX, BY and CZ be the heights relative to the bases BC, AC and AB respectively. We will hide their intersection.

Let (l_1) be the line passing in C and parallel to (AB), (l_2) be the line passing in B and parallel to (AC), and (l_3) be the line passing in A and parallel to (BC). Let D, E and F be the points of intersection of (l_1) and (l_2), (l_1) and (l_3), (l_2) and (l_3) respectively.

1. Prove that $ABCE$, $ACBF$ and $ACDB$ are parallelograms.
2. Prove that A, B and C are the midpoints of $[EF]$, $[FD]$ and $[DE]$ respectively.
3. Deduce that (AX), (BY) and (CZ) are the perpendicular bisectors of $[EF]$, $[FD]$ and $[DE]$ respectively.
4. Deduce that the 3 heights of triangle ABC meet in a common point.



7. Attractive Induced Formula

1. Using your calculator find the values of: $\frac{13}{75}$, $\frac{1313}{7575}$, $\frac{131313}{757575}$.
2. What can you conclude?

3. Can you give a 3-digit example with the same previous feature?
4. Show that: "If $\frac{a}{b}=\frac{c}{d}$ then $\frac{a}{b}=\frac{a+c}{b+d}$". [Note that this part is asked for grade 8; if the problem is given to grade 7, it is skipped.]
5. Use the following property to give the reason behind the above equalities?

$$\text{“If } \frac{a}{b}=\frac{c}{d} \text{ then } \frac{a}{b}=\frac{a+c}{b+d}\text{”}$$

6. Can we deduce that $\frac{25}{471}=\frac{2525}{471471}$? Why? (try to use the formula)
7. Apply the strategy on the 3-digit-number example.

Appendix B

Q1 According to your teaching experience, does this problem make use of any mathematical concepts that students of this grade do not know or are not aware of?

Q2 According to your teaching experience, do you find any part of this problem hard for students to solve?

Q3 According to your teaching experience, do you find the solution of any part of this problem hard for students to understand?

Q4 In your opinion, do you think that this problem triggers students' curiosity and help them see that mathematical properties and formulas are linked together?

Q5 In your opinion, does this problem represent a good example of our notation "the unexpected links"?

Q6 Any remarks?

Appendix C

Questionnaire

Choose <u>one</u> answer only:

1. I understood:
   a. The whole problem.
   b. Nothing in the problem.
   c. Other (some parts).

2. The problem was interesting:
    a. Yes.
    b. No.

References

Aigner, M., & Ziegler, G. M. (2010). *Proofs from the Book* (fourth ed.). Verlag Berlin Heidelberg: Springer.

Booden, M. A. (2004). *The Creative Mind: myths and mechanisms* (2nd ed.). London and NewYork: Routledge - Taylor & Francis.

Brinkmann, A., & Sriraman, B. (2009). *Aesthetics and Creativity: An exploration of the relationships between the constructs.* Charlotte NC: Information Age Publishing (2009).

Chamberlin, S. A., & Moon, S. M. (2005). Model-Eliciting Activities as a Tool to Develop and Identify Creatively Gifted Mathematicians. *The Journal of Secondary Gifted Education, XVII*(1), 37-47.

Connes, A. (n.d.). *A View of Mathematics.* Retrieved from www.alainconnes.org/docs/maths.pdf

De la Vallée Poussin, C. J. (1896). Recherches analytiques sur la théorie des nombres premiers. *Ann. Soc. Sci. Bruxelles, 20*, 183–256.

Fitzpatrick, R. (2007). *Euclid's Elements of Geometry. The Greek text of J. L. Heiberg (1883-1885) from Euclidis Elementa, edidit et Latine interpretatus est I.L. Heiberg, in aedibus. B.G. Teubneri, 1883-1885.* (R. Fitzpatrick, Ed., & R. Fitzpatrick, Trans.) Richard Fitzpatrick, 2007.

Hadamard, J. (1893). Etude sur les proprietés des fonctions entiéres et en particulier d'une fonction considérée par Riemann. *Journal de mathématiques pures et appliquées 4e série, 9*, 171-216.

Hadamard, J. (1896). Sur la distribution des zéros de la fonction ζ(s) et ses conséquences arithmétiques. *Bulletin de la S. M. F., 24*, 199-220.

Hardy, G. H. (1940). *A Mathematicians Appology.* First electronic edition, version 1.0, March 2005. The University of Alberta Mathematical Sciences Society.

Haylock, D. (1997). Recognizing Mathematical Creativity in School Children. *ZDM, 29*(3), 68-74.

Kleiner, I. (2012). *Excursions in the History of Mathematics.* Birkhauser, Springer Science + Buisness Media, LLC 2012.

Leikin, R. (2009). Exploring mathematical creativity using multiple solution tasks. In R. Leikin, A. Berman, B. Koichu, R. Leikin, A. Berman, & B. Koichu (Eds.), *Creativity in Mathematics and the Education of Gifted Students* (pp. 129-145). Rotterdam, The Netherlands: Sense Publishers.

Liljedahl, P., & Sriraman, B. (2006, March). Musings on Mathematical Creativity. *For the Learning of Mathematics, 26*(1), 17-19.

Mann, E. L. (2005). *Mathematical Creativity and School Mathematics: Indicators of Mathematical Creativity in Middle School Students.* PhD Dissertation, University of Connecticut.

Nadjafikhaha, M., Yaftianb, N., & Bakhshalizadehc, S. (2012). Mathematical creativity: some definitions and characteristics. *Procedia - Social and Behavioral Sciences, 31*, 285 – 291.

Rosen, F. (1831). *The algebra of Mohmmed ben Musa.* (F. Rosen, Ed., & F. Rosen, Trans.) Oriental Translation Fund.

Silver, E. A. (1997). Fostering Creativity through Instruction Rich in Mathematical Problem Solving and Problem Posing. *ZDM, 29*(3), 75-80.

Sinclair, N. (2009). Aesthetics as a liberating force in mathematics education? *ZDM Mathematics Education, 41*, 45–60.

Sriraman, B. (2004). The Characteristics of Mathematical Creativity. *The Mathematics Educator, 14*, 19-34.

Sriraman, B., Haavold, P., & Lee, K. (2013). Mathematical creativity and giftedness: a commentary on and review of theory, new operational views, and ways forward. *ZDM Mathematics Education, 45*, 215–225.

Stillwell, J. (2010). *Mathematics and its History* (third ed.). Springer Science+Buisness Media, LLC 2010.

Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing Creativity: A Guide for Educators.* University of Connecticut. The national research center on the gifted and talented.

Verhulst, F. (2012, September 3). An interview with Henri Poincaré: Mathematics is the art of giving the same name to different things.

Weisberg, R. W. (1988). Problem solving and creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 148-176). New York: Cambridge University Press.

Whitcombe, A. (1988, March). Mathematics: creativity, imagination, beauty. *Mathematics in School, 17*(2), 13-15.

A DIALECTICAL INVARIANT FOR RESEARCH IN MATHEMATICS EDUCATION

Mauro García Pupo[1]
Universidad Antonio Nariño
Colombia

Juan E. Nápoles Valdes[2]
Universidad Nacional del Nordeste
Universidad Tecnológica Nacional
Argentina

Abstract

Many current problems in research in mathematics education emerge from pairs of contradictory dialectical categories. In effect, these pairs characterize the problems. When an epistemological study is made to determine the object of research in which a problem is immersed, it is possible to find essential pairs of dialectical categories that become more profound and thus provide enough elements for the determination of appropriate didactic actions to solve the problem under research.

2010 Mathematics Subject Classification: 97D20

Keywords and phrases: dialectical categories, object of investigation, research problems in mathematics education.

Introduction

A central issue in the discipline is the answer to the question: what is mathematics education? *"Many answers to these questions have traditionally been, and continue to be, advanced. Standard reasons include the need to produce another generation of scholars to continue developing the discipline of mathematics, the supply of a cadre of scientists and others such as engineers who need strong mathematical competence, as training in logical thinking and problem solving, as exposure to what is as much a part of cultural heritage as literature or music. All of these, and more, are valid, but a deeper analysis is required"*[3].

Carlos Vasco models mathematics education as an octagon in which mathematics (which he classifies as research, scholars and daily life) is located on the inside and on the outside, forming the sides, are eight disciplines: philosophy, logic, computer science, linguistics, neurology, psychology, anthropology, history of mathematics and epistemology[4].

We believe that mathematics education integrates dissimilar disciplines as they are represented in Figure 1 which clearly illustrates the complexity on which we base the considerations that follow.

[1] mauro@uan.edu.co
[2] jnapoles@exa.unne.edu.ar and jnapoles@frre.utn.edu.ar
[3] Greer, B. What is mathematics education for? Portland State University, Portland, U.S.A.
[4] Vasco U.,C.E.(1994). La educación matemática: una disciplina en formación, *Matemática Enseñanza Universitaria*, Vol III, Nro 2, 59-76.

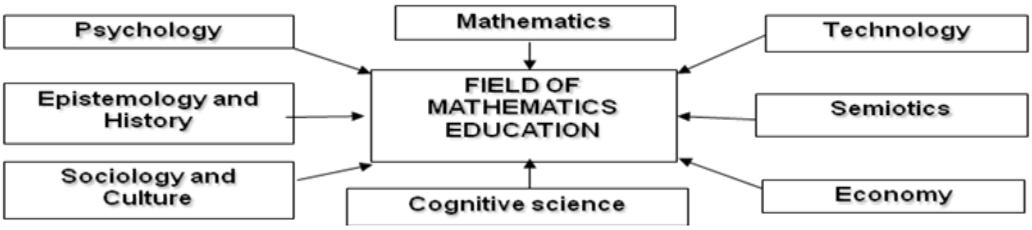


Figure 1

Other important reflections are contained in the following sentences: *"So mathematics education is fashioned to provide appropriate mathematical knowledge, understanding, and skills to diverse student populations"*[5]*. In this sense we can say that mathematics education is a special type of teaching, engineering in the sense that it is the personalization of the basic mathematical principles to meet the needs of teachers and students*[6]*, some even claim that it is a branch of applied mathematics*[7].

Mathematics education today is a science that at a glance reveals two conflicting features: on the one hand theoretical sublimation and on the other a multiplicity of practical problems that must be resolved by the teacher in the classroom. Of course, this two-faced trait manifests itself in research carried out in different regions of the world. In some cases "theoretical centers" immersed in general issues are at the center of the research being done, and in others there are institutions dealing with the "real" problems that teachers face in the classroom.

A principal point must be made clear: what is a research in mathematics education?

Dialectic theory delimits knowledge formation as an active, complex, ongoing process of organizing and reorganizing conceptual structures rather than an accumulation of fixed truths. Furthermore, in dialectic theory contradiction assumes a central role in the process of change and reorganization that the theory presumes to explain. Whether in the field of cognitive development or in the broader realm of psychology, a dialectical view also assumes that developmental processes are socially and culturally shaped and defined, and that concepts and meanings—whether mathematical or not - evolve in an emergent process of what Vygotsky[8] and Leontiev[9] called a collective activity system. "The latter is understood to operate through the emergence of cognitive conflict within the conceptual system, leading to the ongoing resolution of that conflict in a dialectical manner- which is to say through the

[5] Bass, H. (2005). Mathematics, mathematicians, and mathematics education, *Bulletin Amer. Math. Society*, 42, 417-430 y Ferrini-Mundy, J. and Findell, B. (2001): The mathematics education of prospective teachers of secondary school mathematics: old assumptions, new challenges, *in CUPM Discussion Papers about Mathematics and the Mathematical Sciences in 2010: What Should Students Know?* Washington DC: Mathematical Association of America.

[6] Wu, H. (2006). How mathematicians can contribute to K-12 mathematics education, *Proceedings of International Congress of Mathematicians*, Madrid 2006, Volume III, European Mathematical Society, Zürich, 2006, 1676-1688 (http://math.berkeley.edu/~wu/ICMtalk.pdf) and "Access and Opportunities to Learn Are Not Accidents: Engineering Mathematical Progress in Your School" by William F. Tate, which is available at http://www.serve.org/downloads/publications/Access_AndOpportunities.pdf. The concept of mathematics education as mathematical engineering also sheds some light on Shulman's concept of *pedagogical content knowledge* (see Shulman, L. (1986). Those who understand: Knowledge growth in teaching, *Educational Researcher*, 15, 4-14).

[7] Bass, H. (2005). Mathematics, mathematicians, and mathematics education, *Bulletin Amer. Math. Society*, 42, 417-430 y Ferrini-Mundy, J. and Findell, B. (2001). The mathematics education of prospective teachers of secondary school mathematics: old assumptions, new challenges, *in CUPM Discussion Papers about Mathematics and the Mathematical Sciences in 2010: What Should Students Know?* Washington DC: Mathematical Association of America.

[8] Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

[9] Leontiev, A. (1981). *Problems of the development of the mind*. Moscow: Progress.

recognition and articulation of contradictions and inconsistencies, and their mediation in the context of a collective activity"[10].

The teleology of the processes of structural cognitive development, as defined by the major western theorist of cognitive change, Piaget, is understood not only as a continuous movement from "no balance" towards reequilibration, but as progressively directed movement towards "increasing equilibrations," which necessarily require a correspondingly higher organization of cognitive structure. Although it was never affirmed by Piaget, several theorists[11] understand his psychological theory of cognitive development to be fundamentally dialectical. We can identify many parallels between Piaget and the other major theorist of the twentieth century, Vygotsky, at least on the level of the basic, conceptual mechanisms of cognitive development. Vygotsky was in fact an avowed dialectician, who clearly saw cognitive development as *"…a dialectical process, a process in which the transition from one stage to the next occurs not through evolution, but through revolution"*[12].

Ever since Imre Lakatos presented a dialectical view of the development of mathematical knowledge in his *Proofs and Refutations* (1976), the idea of carrying out dialectical processes in the mathematics classroom has often attracted the attention of mathematics educators[13].

On the other hand, the larger scale view of activity provided by this perspective considers learning in terms of fundamental qualitative changes in an activity system as a whole, a process that Engeström calls *expansive learning*. This occurs as a result of deliberate efforts of participants over time to solve inherent conflicts and contradictions that are a part of any activity system. Engeström's theorization does not provide an explicit direction for understanding the place of mathematics within a given activity, nor does it provide details related to the learning process of individuals[14].

Based on his reading of Vygotsky's semiotics, Leontiev's activity theory, and the more recent work of Felix Mikhailov and Evald Ilyenkov, Radford has developed the *Theory of Objectivization* specifically for unpacking nuances and processes of mathematics activity and learning of individuals from a cultural-semiotic activity perspective[15]. In contrast to Engeström, Radford's work focuses on specific aspects of the consciousness, learning and being of individuals as well as on the semiotic and social dimensions of mathematics from an activity perspective. Radford's concept of objectivization is a refinement of Vygotsky's

---

[10] Kennedy, N. S. (2006). Conceptual change as dialectical transformation, in Novotná, J., Moraová, H., Krátká, M. & Stehlíková, N. (Eds.). *Proceedings 30th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 5, pp. 193-200. Prague: PME.

[11] Kitchener, R. (1986). *Piaget's theory of knowledge: Genetic epistemology & scientific reason*. New Haven, MA: Yale University Press.

[12] Vygotsky, [The problem of age], cited in El'konin, D. (1977). Towards the problem of stages in the mental development of the child. In M. Cole (Ed.), *Soviet developmental psychology* (pp. 539-563). New York: Sharpe. p. 542.

[13] For a critical review, see Gila Hanna: The Ongoing Value of Proof, in Luis Puig and A. Gutierrez (Eds) *Proceedings of the International Group for the Psychology of Mathematics Education*, Valencia, Spain, Vol I., 21-34.

[14] Engeström, Y. (2001). Expansive learning at work: toward an activity theoretical reconceptualization, *Journal of Education and Work*, *14*(1), 133-156 and Engeström, Y. (2008).*From teams to knots: Activity-theoretical studies of collaboration and learning at work*, New York: Cambridge University Press.

[15] Radford, L. (2006). Elements of a cultural theory of objectification, *Revista Latinoamericana de Investigación en Matemática Educativa, Special issue on semiotics, culture and mathematical thinking*, pp. 103-129, and Radford, L. (2007). Towards a cultural theory of learning, in Pitta-Pantazi, D., & Philippou, G. (Eds.). *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education (CERME – 5)*, pp. 1782-1797. Larnaca, Cyprus, CD-ROM, ISBN - 978-9963-671-25-0 and Radford, L. (2008). The ethics of being and knowing: Towards a cultural theory of learning, in L. Radford, G Schubring, & F. Seeger (Eds.), *Semiotics in mathematics education: Epistemology, history, classroom and culture* (215-234). Rotterdam: Sense Publishers.

notion of internalization which emphasizes the dialectical relationship between the subject and the cultural object being attended to.

There are other theories that use the dialectical approach from other perspectives. For instance, Dubinsky theorizes that mathematical objects are constructed by reflective abstraction in a dialectic sequence **APOS**, beginning with **A**ctions that are perceived as external, interiorized into internal **P**rocesses, encapsulated as mental **O**bjects developed within a coherent mathematical schema[16]. Drossos, on the other hand, uses the opposite dialectical *assimilation* and *accommodation* when he talks about *adaptation* in the process of cognition[17].

In this paper we show how contradictions, immersed in a large part of the problems related to mathematics education, are a guiding source for didactics modeling that can lead to the solution of the problem researched and to the production of results.

METHODOLOGY AND RESULTS

We will focus our attention on two features of research in education in general and in mathematics education in particular:

1. Difficulties in the emergence of new knowledge.
2. The role played by the dialectical method in interrelating the research subject and the object of research.

How does a researcher discover a particular object of research?

The formulation of a research problem in mathematics education is frequently related to difficulties in the teaching-learning process of a mathematical topic at some level of education. It proceeds consciously or unconsciously through a process of abstraction of contradictory dialectical categories[18]. We consider that a contradiction exists only if it has a witness. That means that a contradiction does not exist by itself, but only with reference to a cognitive system[19]. According to Piaget (1974, p. 161), the awareness of a contradiction is only possible at the level at which the subject becomes able to overcome it[20]. We consider that most of the problems in the teaching of mathematics are characterized by a contradiction between dialectical pairs of students' knowledge and their level of achievement.

In the theory of situations, the term 'dialectic' refers to the method used by a cognitive system (teacher, student) to manage the contradictions between its expectations concerning the output from the system it attempts to control (the student-milieu system, the teacher-milieu, respectively) and the feedback. Feedback is communication of information. The process of dialectic turns this information into knowledge: out of the contradiction, something positive is attained that explains the contradiction and generates ways of avoiding it in the future.

[16] Dubinsky. E. & MacDonald, M. A. (2001). APOS: A Constructivist Theory of Learning in Undergraduate Mathematics Education Research, in D. Holton *et al.* (Eds.), *The Teaching and Learning of Mathematics at University Level: An ICMI Study*, Dordrecht: Kluwer, 273-280 http://www.math.kent.edu/~edd/ICMIPaper.pdf

[17] Drossos, C. A.(1987). Cognition, Mathematics and Synthetic Reasoning, *General Seminar of Mathematics*, Department of Mathematics, University of Patras, Greece vol. 13, 107-151.

[18] Godino, J. D.; C. Batanero and V Font (2007).The onto-semiotic approach to research in mathematics education, *ZDM Mathematics Education*, 39:127–135.

[19] Cf. Grize, J.B.; Piéraut-le Bonniec (1983). La Contradiction, Paris: PUF and Balacheff, N.(1991).Treatment of refutations: aspects of the complexity of a constructive its approach to mathematics learning, E. von Glasersfeld (ed.) (1991). Radical Constructivism in Mathematics Education, 89-110, Kluwer Academic Publishers.

[20] Piaget, J. (1974*). Les Relations entre Affirmations et Négations*, Recherches sur la contradiction, Vol. 2, Paris: PUF, p. 161.

Resolution of contradictions in each case (either in the *situation of action* or in the *situation of formulation*) brings some positive new knowledge about the situations: a better way of expressing one's ideas or an improved strategy[21].

Usually this situation emerges in the classroom, almost always far removed from the possibility of solving it by means of a scientific process. Ignoring the difficulties in learning, it can be characterized and didactical solutions can be sought focusing on the identification of the object of research in the process of epistemologization as described as follows.

Primary contradiction → Research problem → Object of research

This triple indicates the path starting from a primary contradiction (which is evident), and may be referred to as an external contradiction. When this path for seeking scientific knowledge in mathematics education is assumed, then it is possible to find, perhaps through a series of steps, better refinements towards possible resolutions to the problem being researched, i.e., a succession of contradictions {$\mathbf{C_n}$}, from which we have:

External contradiction = $C_0 \to C_1 \to \ldots \to C_n$ = Fundamental contradiction

From the problematic situation that has been detected, we build the research design, and with its help a better understanding of the research problem, closely related to the object of study and the proposed objective. As refinements are achieved, we will find a better approach to an **object to study** or valid **object of research** and, of course, toward a **field of a** relevant **action**.

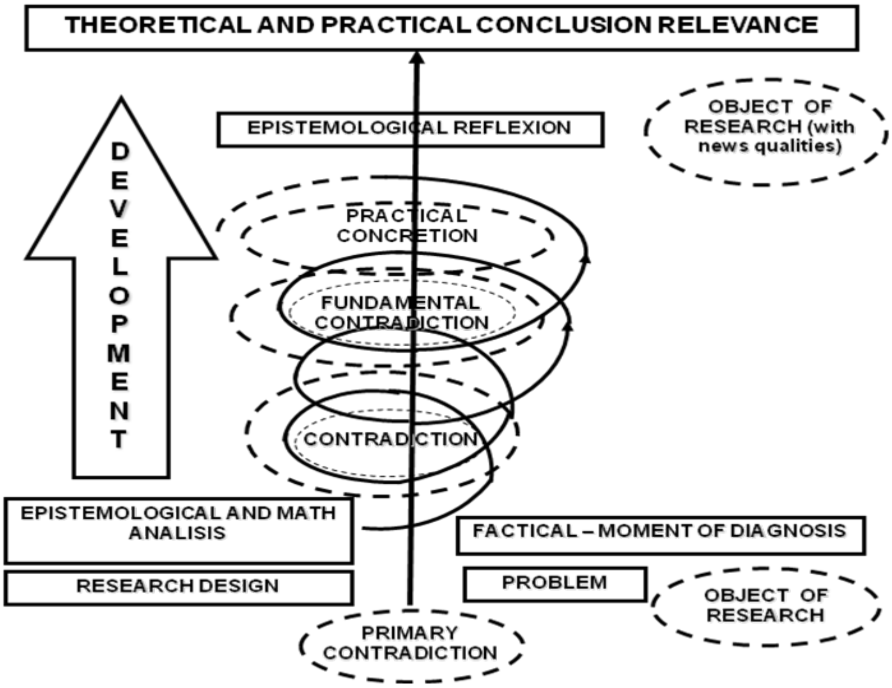


Figure 2

Throughout this article the system of contradictory pairs will appear repeatedly because of their invariant nature, even after conclusion of the investigation, that is, the new object of research that is defined above must contain, in a natural way, a pair of new dialectical contradictions, qualitatively higher.

---

21 More details in Sierpinska, A. Theory of situations as a means to overcome the 'procedures vs understanding' dilemma in mathematics teaching, MATH 645: Theory of Situations/Lecture 2 (in http://annasierpinska.wkrib.com/pdf/TDSLecture%202.pdf consulted on February 3th of 2012).

In Figure 3 we can see the actions of a teacher in the process of didactic transposition, as a simple case of the previous figure. In fact, a path from the concrete to the abstract is shown; both categories are dialectically connected through the activities of human beings whose actions are intrinsically related to the activity setting which represents a multi-faceted, yet organized, whole. Abstraction is a process of making sense of such concrete situations by discovering new meanings in order to establish connections amongst the different elements of the whole[22].

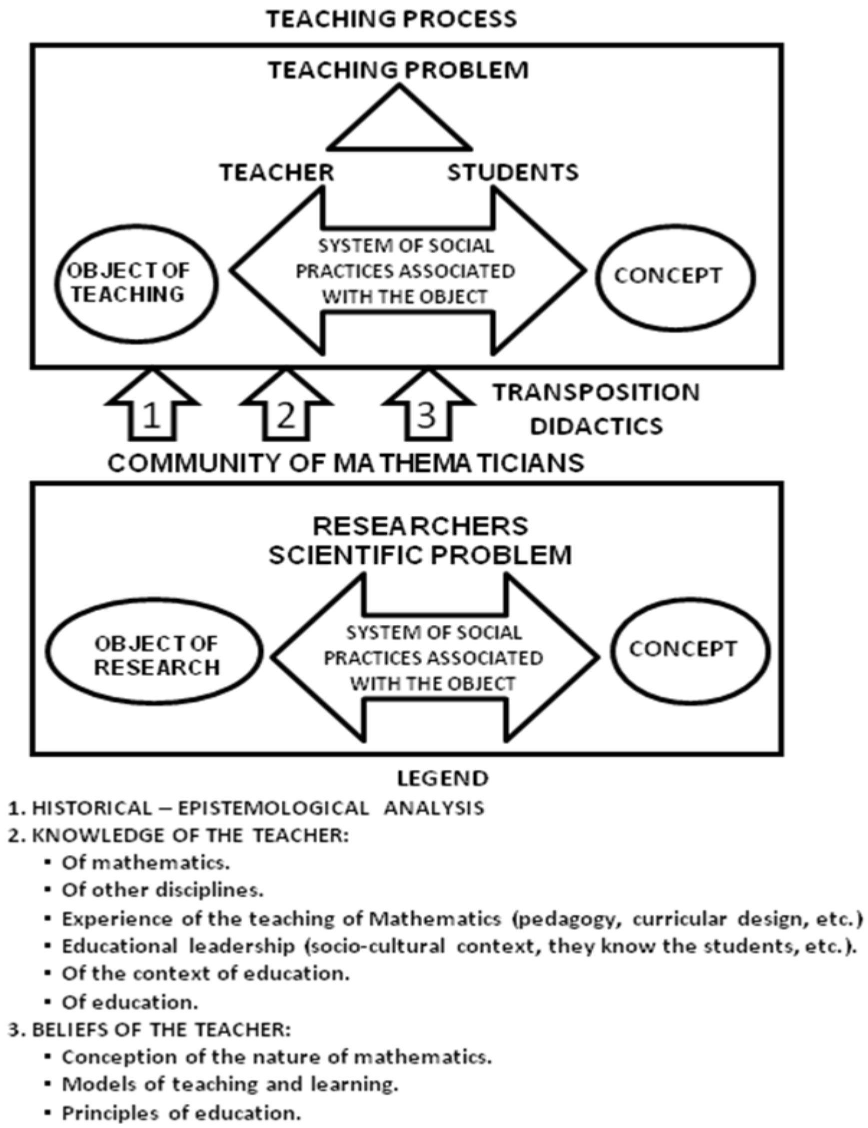


Figure 3

In Figure 4, an example is shown in which we can appreciate a refinement made in a recent research experience in the teaching of geometry. From the external contradiction or **contradiction 0** the process led the researcher to the fundamental contradiction or **contradiction n** in three steps:

[22] Ozmantar, M. F. and J. Monaghan (2007). A Dialectical Approach to the Formation of Mathematical Abstractions, *Math. Education Research J.* Vol. 19, No. 2, 89–112.

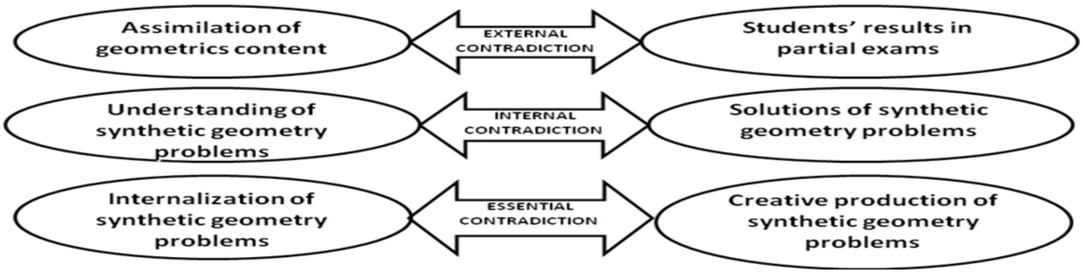


Figure 4

On a more general level, Asano notices several differences between dialectics and didactics of mathematics, on the basis that they are concerned with different kinds of objects, intermediates and forms; however, his observations principally treat differences in methods and it is important to keep in mind a long standing appreciation in this regard. Plato talks about mathematics and dialectics in the simile of the line and subsequent pages of Book VI of *The Republic*[23]. Plato's educational priorities also reflected his distinct pedagogy. Challenging the Sophists (who prized rhetoric, believed in ethical and epistemological relativism, and claimed to teach "excellence"), Plato argued that training in "excellence" was meaningless without content. Plato doubted whether a standard method of teaching existed for all subjects, and he argued that morally neutral education would corrupt most citizens. He preferred the dialectical method over the Sophists' rhetorical pedagogy.

Before continuing, let us look at some of the aspects that determine quality in research in mathematics education. A result of quality in research in mathematics education can be characterized as follows:

1. It shows new relationships and regularities that the researcher reveals in the process of resolving the problem.
2. Along with scientific background results, it makes a difference in the resolution of similar problems.
3. It works as a systemic feature of transformation of the process being modeled.
4. The model supports the theoretical contributions of the thesis or research result as well as the essence of the text being written. Some contributions can be:
    - Problem solving.
    - Mathematical technique
    - Mathematical education theory
    - Mathematical exposition
    - Mathematical pedagogy
    - Mathematical vision and visualization
    - Rigorous mathematics
    - Beautiful mathematics
    - Elegant mathematics
    - Creative mathematics
    - Intuitive mathematics

[23] Cf. Asano, K. (1993). Degrees of Reality in Plato: Part I." Aichi (Philosophy) Vol. 10: 131–118; (1994). Degrees of Reality in Plato: Part II. The Hannan Ronshu (*Journal of Hannan University*) *Humanities & Natural Science*, Vol. 30, No. 2 (Sep.): 17–34; (1996). Two Arguments for Forms in Plato: Conflicting Appearances and One over Many. The Hannan Ronshu (*Journal of Hannan University*) Humanities & Natural Science, Vol. 32, No. 1 (June): 79–96; (1997A). The Simile of the Line in Plato's Republic VI. Sapientia (*The Eichi University Review*) No. 31: 207–34; (1997B). A Study of Plato's Metaphysics in the Republic." Ph.D. diss., University of Texas, Austin and (1998). Mathematics and Dialectic in Plato's Republic, Sapientia (32):117-142.

These results, seen as new knowledge for science, are identified by the educational community of the sciences as a **theoretical contribution** to research which may or may not lead to the obtention of a scientific degree.

Research reports in the area of mathematics education have revealed argumentative inadequacies of researchers when they attempt to prove the scientific novelty of their theoretical contribution. What this paper proposes is a general procedure to promote understanding for the modeling of the theoretical scientific contribution of research in the pedagogical sciences in general and in mathematics education in particular[24].

Conception of theoretical contribution in research in mathematics education.

The results of research should contain contributions on the theoretical and practical levels. Contributions on the theoretical level are embodied in models, definitions, concepts, characterizations, revelation of phenomena of a periodic nature, among others. Practical contributions include methodologies, strategies, techniques, procedures, among others, and they work as tools for the implementation of the theory. The main contribution is the theoretical conception or theoretical contribution that underpins the research.

Both the identification of a problem and its resolution are supported in the dialectical approach to knowledge. The formulation of the scientific problem is the expression of a contradiction between objects and/or phenomena contrasting the current state and the desired state, which we have previously called an external contradiction. This external contradiction is revealed in the initial diagnosis of the object based on empirical data obtained by means of an exploration of an actual situation and a bibliographic review that allows the construction of theoretical foundations for the problem as revealed in the state of the art. As a result, the object of study and field of action (unit of analysis) in an initial stage of the research process may be described.

Ways of promoting the dialectical method of knowledge for the construction of a theoretical conception which is the basis for the resolution of a scientific problem

Here we will see the methodological value of the dialectical method. The dialectical approach allows the analysis of the most essential aspects of the object, analysis which consists of determining those contradictory elements which are present in it. The opposites are mutually exclusive aspects of the object that at the same time question each other. The mutual relationship between the opposites constitutes a contradiction.

The contradiction plays its role as source of the movement and the development of the object and the phenomenon. Contradictions of the object or phenomenon with other objects or phenomena are considered external contradictions. Internal contradictions are formed between opposing aspects of the object itself or given phenomenon in themselves. When is not possible to refine an internal contradiction, then it plays the decisive role in the development of the object or phenomenon being in this case the fundamental contradiction.

Our use of dialectic follows ancient Greek thought. Unlike the more recent Hegelian use of the term that anticipates a synthesis of opposites, we want to revitalize an earlier sense of dialectic that predates Plato and views dialogue, discourse and dispute themselves as deepening our understanding of the world[25]. Dialectic is a kind of juxtaposition of ideas,

[24] Concepción, R. y Rodríguez, F. (2005). Un procedimiento para elaborar el aporte teórico de la tesis de doctorado en Ciencias Pedagógicas basado en el enfoque sistémico-estructural, Universidad de Holguín, Cuba.

[25] Parmenides' (510 BC). Foundational poem is seen as a starting point for the ongoing development of the idea of dialectic: *"There is need for you to learn all things ... both the unshaken heart of persuasive Truth and the opinions of mortals, in which there is no true reliance ... that the things that appear must genuinely be, being always, indeed, all things"* in Diels, H. & Kranz, W. (1951). *Die*

often literally a debate, rather than a resolution or synthesis. Understandings emerge by means of holding in creative tension ideas that can even seem paradoxical.

We assume, accordingly, that the resolution of the research problem with a dialectical approach should reveal the existence of contradictory pairs of elements present in the object of research (fundamental contradiction) and their resolution through a third procedural element, concurrent and simultaneous with the other two, and such that through its introduction it is possible to accelerate the inherently dynamic nature of a dialectical contradiction. This can be achieved throughout the research process[26].

The challenge imposed by the analysis of the object using a dialectical approach is that of specifically determining the primary contradictory pair and of discovering a third element which is also contradictory to the original couple and thus stimulate the transformation of the problem. This analysis permits the characterization of the object of research and of the field of action through a model or theoretical conception, which becomes an instrument of optimization and forms the basis of the proposal or contribution of the research.

Ways of building the theoretical conception or model which constitutes the essential theoretical contribution of the research

The theoretical model or the modeling of the theoretical conception is a construction that the researcher creates or develops starting from his theoretical knowledge of the object of research and the field of action, it is, as all kinds of models are, the idealization (abstraction-realization) the researcher makes in order to transform the process.

The conception of a theoretical contribution is defined as a personal construction of the researcher, product of the abstraction of the object and process that seeks to transform, in which the latter is reproduced in its totality by means of the relationship between contradictory elements that can accelerate the movement and development of that process in a given social historical context.

The realization of a theoretical contribution in itself constitutes the manner of achieving new concrete knowledge as thought through by the researcher. According to Alvarez de Zayas, the representation or theoretical construction can be presented as a theoretical model and its totalizing conception must be achieved in order to conceive this as a system[27]. That is to say, the construction of the model is favored if it is treated with a systemic approach. So it is advisable, following this approach, to build the model of the theoretical contribution based on the general characteristics of systems.

Components of the system: These are the fundamental elements that characterize the model and which are essential to resolve the problem.[28] They must include concepts and categories that the researcher has discovered in order to abstract the object or phenomenon that is modeled. The components are the contradictory pair and the third co-existing element that causes or resolves the fundamental contradiction as well as other elements such as dimensions or variables that permit the understanding of the object or phenomenon to be modeled. The components of the system should acquire their own personality in the object or process to be idealized and are contextualized to the activity in which this contradiction and its resolution take life.

Fragmente der Vorsokratiker. Berlin: Weidmann, translated into English by Kathleen Freeman in her *Ancilla to the Pre-Socratic Philosophers*. Oxford: Basil Blackwell, 1962, p. 246.

[26] Álvarez de Zayas C.(2000). Metodología de la investigación científica. Cómo se modela la investigación científica (in digital format)

[27] Ibidem

[28] Ibidem

Structures and their functional relationships: These provide the framework for interaction and organization among the components of the system, necessary to assure its functions. The structure guides the procedure or mechanism that sparks the activity or process that is modeled. Relationships explain the dynamics of behavior.

Hierarchy: This is the degree of interaction between subsystems.

The process of elaboration or conception of the theoretical contribution requires a scientific abstraction and represents the essential qualitative jump that a researcher must make in order to make contributions to the science being researched.

Personal experience and the meta cognitive diagnosis of how researchers operate in the elaboration of a theoretical conception in research in the pedagogical sciences, reveal that such research unfolds like a process of successive scientific abstractions which are facilitated by means of a procedure that orients the phases and actions of this process and are necessary to attain the objective.

General procedure for the construction of a theoretical contribution.

The procedure that sets out to elaborate a theoretical conception or contribution is supported by the postulates of the theories of the systemic and dialectic approaches to knowledge construction and scientific modeling. It is important to keep in mind that the application of this procedure in itself does not guarantee arriving at a theoretical contribution if the researcher does not have well defined the establishment of the theory that serves as base for the research process. The procedure is a manner of establishing the components, their relation and structure in the modeling of the theoretical contribution and facilitating the search for arguments that explain it.

The procedure consists of questions, actions and phases. The questions are formulated with a metacognitive intention of orienting the investigator towards the reflective understanding of the actions that are chosen for each phase. The phases integrate the actions in the process of scientific elaboration, in the manner of a generalizing succession of analysis and synthesis.

The elaboration of the theoretical conception should demonstrate the logic of the scientific reasoning followed by the researcher in its construction and passes through four phases or moments.

Phases of the construction of a theoretical conception

First phase: Determination of the process or activity object of transformation. This initial phase is crucial for research, because it is when a researcher identifies the process that he or she intends to model, something that requires a profound theoretical preparation on the object and the field of research. The meaning of this process or activity for the research to be carried out is conceptualized.[29].

Second phase: Determination of elements that characterize the process or activity object of transformation and are essential to resolve the problem (components of the model).

Taking into account that the theoretical conception will be modeled with a systemic approach, it is necessary to determine all the elements that make up the model without omitting:

1. The process to be modeled
2. The dimensions of the process that is to be modeled
3. The contradictory pair (fundamental contradiction) present in the object

[29] Observations on the use of internet and the electronic library are interesting in Barry, C. A. (1997). Information skills for an electronic world: training doctoral research students, en *Journal of Information Science*, 23 (3), 225-238.

4. The third element of procedural nature or means for expediting the resolution of the contradiction
5. The context in which the fundamental contradiction occurs and is resolved

Third phase: Organization of the structure of the theoretical conception (structure of the model)

When the elements of the model have been determined, functional relationships and the hierarchy of the system must be analyzed. For the modeling, the components may denote categories or very short phrases linked in such a way that they show the dynamic and the feedback of the system.

Fourth phase: Explanation of the process or activity that is to be modeled (dynamic)

This phase presents the arguments that the new theoretical concepts supply in support of the transformation of practice, although the model must "speak for itself". As all systems they should generate a higher systemic quality that does not belong of any element in particular; the interrelationships between components must be such that if they affect one of them, they will affect the whole system and, in consequence, will not develop new and higher characteristics and qualities. For example, see a didactical model in Figure 5.

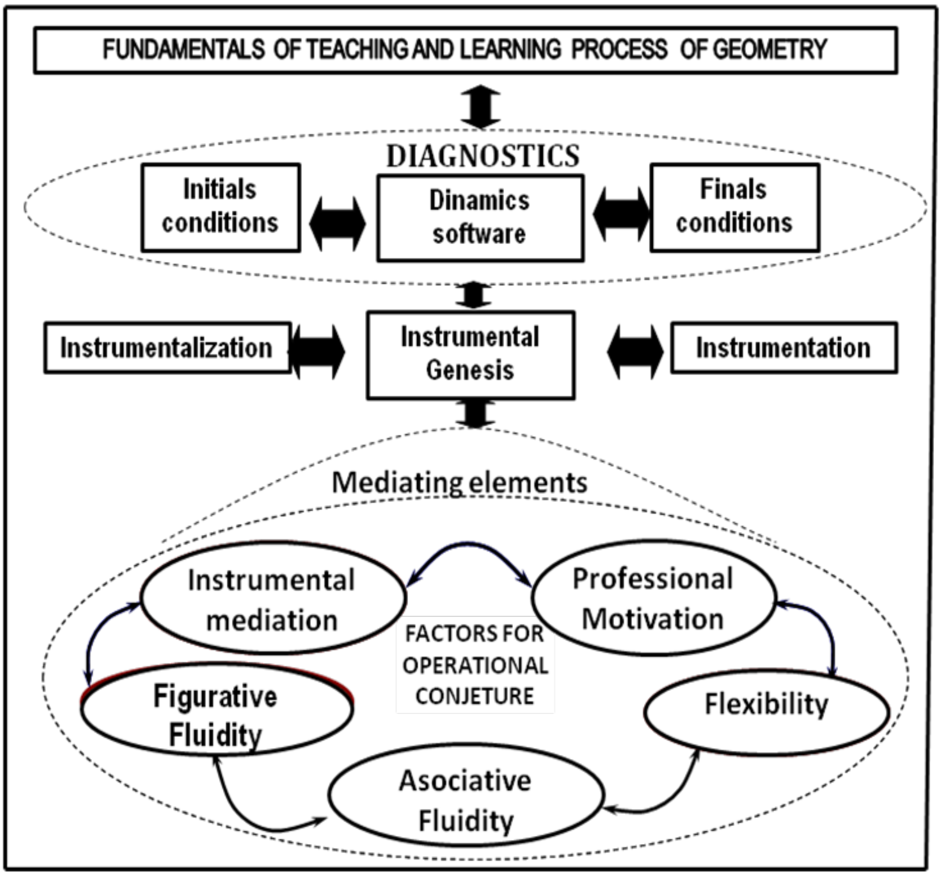


Figure 5: A didactical model[30]

Indications for the elaboration of a heuristic procedure of introduction to practice.

[30] Wilson, C. (2012). Operational conjecture. A didactical model for teaching and learning of geometry in engineering's careers. Thesis in option to the Scientifics grade of Mathematics Education. University of Holguin. (the first author was advisor himself)

In the structure of the procedure, first the questions are formulated; they lead to actions and finally the phases or stages emerge.

The questions are the most dynamic part of the procedure, they give way to a situation in which each researcher performs actions depending on his/her possibilities and needs without becoming an algorithm composed of linear sequences, but rather a process of construction of the scientific text and context that allows the researcher to discover some things and perhaps to return to previous phases to reconsider, or even to find out that phases may overlap.

Finally a pre experiment or a case study must be carried out in which the procedure is introduced into practice, and then interpret or verify the results according to the research paradigm used.

The procedure described should allow the elaboration of an argument that validates the proposed didactical model and, most importantly, the research methodology used to extract the theoretical and practical contributions that raise the object of research to a higher qualitative level. Finally all the results should constitute an argument for the **scientific novelty** that resolved a **gap** and therefore contributed to the **epistemological development** of the research. (See Figure 6 referent to our geometric example.)

Stages of the generation of conjectures	**Actions for the generation of conjectures**	**Guiding questions for the generation of conjectures**	**Resources or heuristic means used**
Construction of the basic elements	**Begin with the construction of basic elements**	**1. What elements should make up the figure or geometric locus?**	**HP: Related to the elements of the figure** **HR: Based on representation of the figure used to analyze it**
	Determine what figures or elements can be generated.	**2. How can I represent the desired figure or geometric locus?**	
	Identify the necessity of each basic element.	**3. Are the basic elements used necessary and sufficient?**	
Movements (transformations) of the elements to initiate a search	**Identify the relations between the basic elements constructed and how they influence one another**	**1. How are the components of the system related to one another?**	**HS: Consists of identifying the basic elements that have been constructed and of observing their behavior using software** **HS: The movement of the mathematical objects is part of the strategy** **HS: Execution of a plan for obtaining a possible solution**
	Move each basic component in coherent manner.	**2. What possible movements can be made?**	
	Identify the pertinence of each basic element.	**3. Is each basic element constructed pertinent?**	
Generation of variants of loci or figures	**Identify the actions that can be performed with each basic element**	**1. Which movements generate geometric figures and loci?**	**HP: Compare the movements of the figure using software in the construction of geometric loci.**
	Look for and determine all the variations that can be generated	**2. Have all the variants been identified?**	

Conjecture what geometric results can be obtained	**Determine the possible figures or geometric loci**	**1. What part of the figure or geometric locus is sustained by the initial basic elements? 2. Can new conjectures be obtained, with other basic elements?**	**HR: Understands the problem and explains or supports it based on the initial basic elements. HP: Relates the problem to other problems.**
Check with paper and pencil	**Manually determine the figure or locus**	**1. What level of coincidence does the verification give rise to? 2. What differentiates the static form obtained from the conventional procedure with the dynamic form employed?**	**HS: Analyze the concept solution using pencil and paper and using dynamic software**

Figure 6. HP (Heuristics procedure), HS (Heuristics strategy) and HR (Heuristics rule)

Conclusions

This is an ascent to the concrete -a process of making meaning by establishing connections amongst elements of the whole- and this is precisely what dialectics is. As Douady affirms, the necessity of organizing a critical work on these cultural answers and also as the acknowledgement of their necessary contribution to the *milieus* with which the students interact, the dialectics between *media* and *milieus* play the essential role.

For instance, new properties of the triangle were found when it was regarded, not in itself, but in connection with the circle. Each triangle can be divided in two right triangles; each one of which can be considered as belonging to some circle. Here the sides and angles appear in totally different interrelations, which were revealed to the eyes of the researcher only by this new relationship. This is a dialectic technique, the technique of theoretical thought. The connection between the triangle and the circle can only be seen as an idea that presupposes the possibility of mentally transforming a triangle into a component of the circle, i.e., reduction of one to the other (of the particular to the general). Only with a transformation, a mental reduction of one figure to another could new properties be detected in the triangle which then laid the foundations of what was a new theory. These properties cannot be revealed by "considering" the triangle in itself and the establishment of the connections defined (reduction of the different one) requires thinking through the concepts.

Theoretical systems, in particular mathematical theories, are always changing, and this includes the scientific theories concerning mathematics education. The words of Karl Popper have a special meaning. *Scientific theories are perpetually changing. This is not due to mere chance but might well be expected, according to our characterization of empirical science"*[31]. Remark: The scientific theories in mathematics education, given the dynamic nature of mathematics education in a context mediated by the modern world, shows this constant change mediated by the laws of dialectics.

The theoretical and practical contributions of an investigation constitute two levels of the concrete that are thought about in the scientific activity organized by the researcher. The theoretical conception or theoretical contribution is built on incorporated scientific

[31] Popper, K. (2009). *The Logic of Scientific Discovery*, Rutledge, New York, p. 50 (of the first English edition on 1959)

knowledge which supports the outcome of the research; it is an essential conclusion that contributes to the science.

The progress that has been made in recent decades is nothing less than phenomenal. In little more than one quarter of a century there have been great epistemological changes, accompanied by a flowering of the tools, techniques and theoretical perspectives that supported them. Cognitive science and socio-cultural research in mathematical education have matured and are becoming more robust; fields that at first seemed to be related almost as thesis and antithesis have, over the last decade or so, generated a synthesis that seems even more promising in terms of its ability to help explain questions concerning (mathematical) thinking, teaching and learning. The same can be said for the artificial distinction between quantitative and qualitative methods that becomes less important when formulating central research questions[32].

In conclusion, we suggest that by solving the research problem we deepen the fundamental understanding of learning, which also helps us in the resolution of many practical issues of teaching. If we start paying serious attention to previous issues, the problems of theory and philosophy will be easier to address and resolve.

References

Álvarez de Zayas, C.(2000). Metodología de la investigación científica. Cómo se modela la investigación científica (in digital format)

Barry, C. A. (1997). Information skills for an electronic world: training doctoral research students, *Journal of Information Science*, 23 (3), 225-238.

Bass, H. (2005). Mathematics, mathematicians, and mathematics education, *Bulletin Amer. Math. Society*, 42, 417-430.

Concepción, R. y Rodríguez, F. (2005). Un procedimiento para elaborar el aporte teórico de la tesis de doctorado en Ciencias Pedagógicas basado en el enfoque sistémico-estructural, Universidad de Holguín, Cuba.

Diels, H. & Kranz, W. (1951). *Die Fragmente der Vorsokratiker*. Berlin: Weidmann, translated into English by Kathleen Freeman in her *Ancilla to the Pre-Socratic Philosophers*. Oxford: Basil Blackwell, 1962, p. 246.

Drossos, C. A.(1987). Cognition, Mathematics and Synthetic Reasoning, *General Seminar of Mathematics*, Department of Mathematics, University of Patras, Greece vol. 13, 107-151.

Dubinsky. E. & MacDonald, M. A. (2001). APOS: A Constructivist Theory of Learning in Undergraduate Mathematics Education Research, in D. Holton *et al.* (Eds.), *The Teaching and Learning of Mathematics at University Level: An ICMI Study*, Dordrecht: Kluwer, 273-280 http://www.math.kent.edu/~edd/ICMIPaper.pdf

Engeström, Y. (2001). Expansive learning at work: toward an activity theoretical reconceptualization, *Journal of Education and Work, 14*(1), 133-156

Engeström, Y. (2008). *From teams to knots: Activity-theoretical studies of collaboration and learning at work*, New York: Cambridge University Press.

Ferrini-Mundy, J. and Findell, B. (2001). The mathematics education of prospective teachers of secondary school mathematics: old assumptions, new challenges, *in CUPM Discussion Papers about Mathematics and the Mathematical Sciences in 2010: What Should Students Know?* Washington DC: Mathematical Association of America.

[32] Schoenfeld, A. H. (2007). *Method*, in F. Lester (Ed.), Second handbook of research on mathematics teaching and learning. New York: MacMillan.

Godino, J. D.; C. Batanero and V. Font (2007).The onto-semiotic approach to research in mathematics education, *ZDM Mathematics Education,* 39:127–135.

Kennedy, N. S. (2006). Conceptual change as dialectical transformation, in Novotná, J., Moraová, H., Krátká, M. & Stehlíková, N. (Eds.). *Proceedings 30th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 5, pp. 193-200. Prague: PME.

Kitchener, R. (1986). *Piaget's theory of knowledge: Genetic epistemology & scientific reason*. New Haven, MA: Yale University Press.

Leontiev, A. (1981). *Problems of the development of the mind*. Moscow: Progress.

Ozmantar, M. F. and J. Monaghan (2007). A Dialectical Approach to the Formation of Mathematical Abstractions, *Math. Education Research J.* Vol. 19, No. 2, 89–112.

Piaget, J. (1974). *Les Relations entre Affirmations et Négations*, Recherches sur la contradiction, Vol. 2, Paris: PUF, p. 161.

Popper, K. (2009). *The Logic of Scientific Discovery*, Rutledge, New York, (of the first English edition on 1959)

Radford, L. (2006). Elements of a cultural theory of objectification, *Revista Latinoamericana de Investigación en Matemática Educativa, Special issue on semiotics, culture and mathematical thinking*, pp. 103-129.

Radford, L. (2007). Towards a cultural theory of learning, in Pitta-Pantazi, D., & Philippou, G. (Eds.). *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education (CERME – 5),* pp. 1782-1797.Larnaca, Cyprus, CD-ROM, ISBN - 978-9963-671-25-0

Radford, L. (2008). The ethics of being and knowing: Towards a cultural theory of learning, in L. Radford, G Schubring, & F. Seeger (Eds.), *Semiotics in mathematics education: Epistemology, history, classroom and culture* (215-234). Rotterdam: Sense Publishers.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching, *Educational Researcher*, 15, 4-14).

Vasco, C.E.(1994). La educación matemática: una disciplina en formación, *Matemática Enseñanza Universitaria*, Vol III, Nro 2, 59-76.

Vygotsky, [The problem of age], cited in El'konin, D. (1977). Towards the problem of stages in the mental development of the child. In M. Cole (Ed.), *Soviet developmental psychology* (pp. 539-563). New York: Sharpe. p. 542.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wilson, C. (2012). Operational conjecture. A didactical model for teaching and learning of geometry in engineering's careers. Thesis in option to the Scientifics grade of Mathematics Education. University of Holguin. (the first author was advisor himself).

Wu, H. (2006). How mathematicians can contribute to K-12 mathematics education, *Proceedings of International Congress of Mathematicians*, Madrid 2006, Volume III, European Mathematical Society, Zürich, 2006, 1676-1688.